



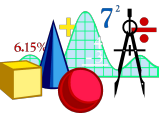
Линейная регрессия

**Стат. методы в
психологии
(Радчи́кова Н.П.)**



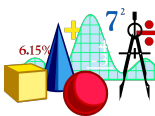
Цели

- ✓ **Зачем проводить регрессионный анализ**
- ✓ **Как проводить регрессионный анализ**
- ✓ **Как интерпретировать результаты регрессионного анализа**



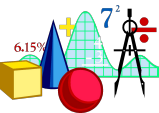


Регрессионный анализ





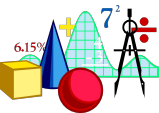
Регрессионный анализ служит для определения вида связи между переменными и дает возможность для прогнозирования значения одной (зависимой) переменной, отталкиваясь от значений других (независимых) переменных.





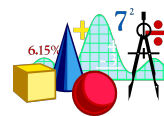
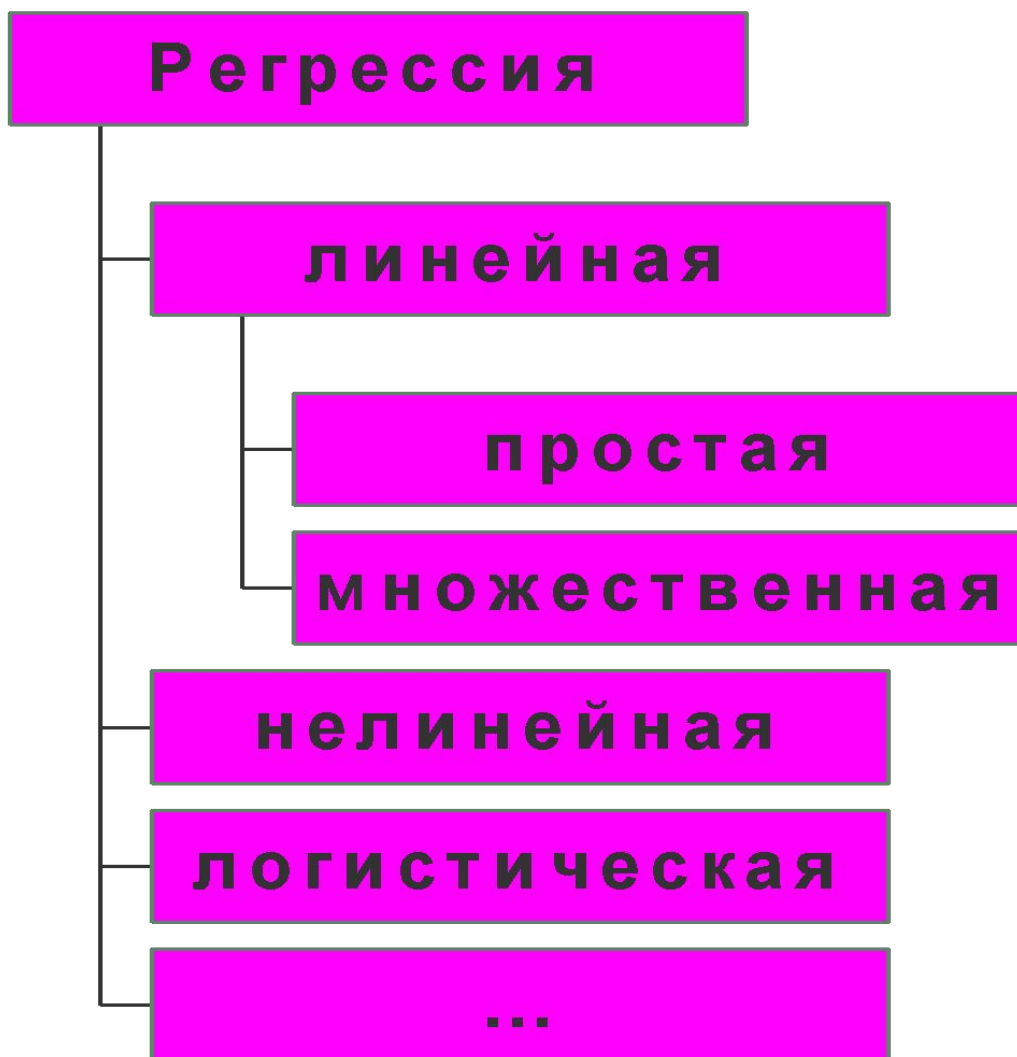
«Регрессионный анализ является мощным средством прогноза. Экономисты, которые им пользовались, успешно предсказали 10 кризисов из 2-х последних»

Материалы Интернета



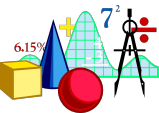
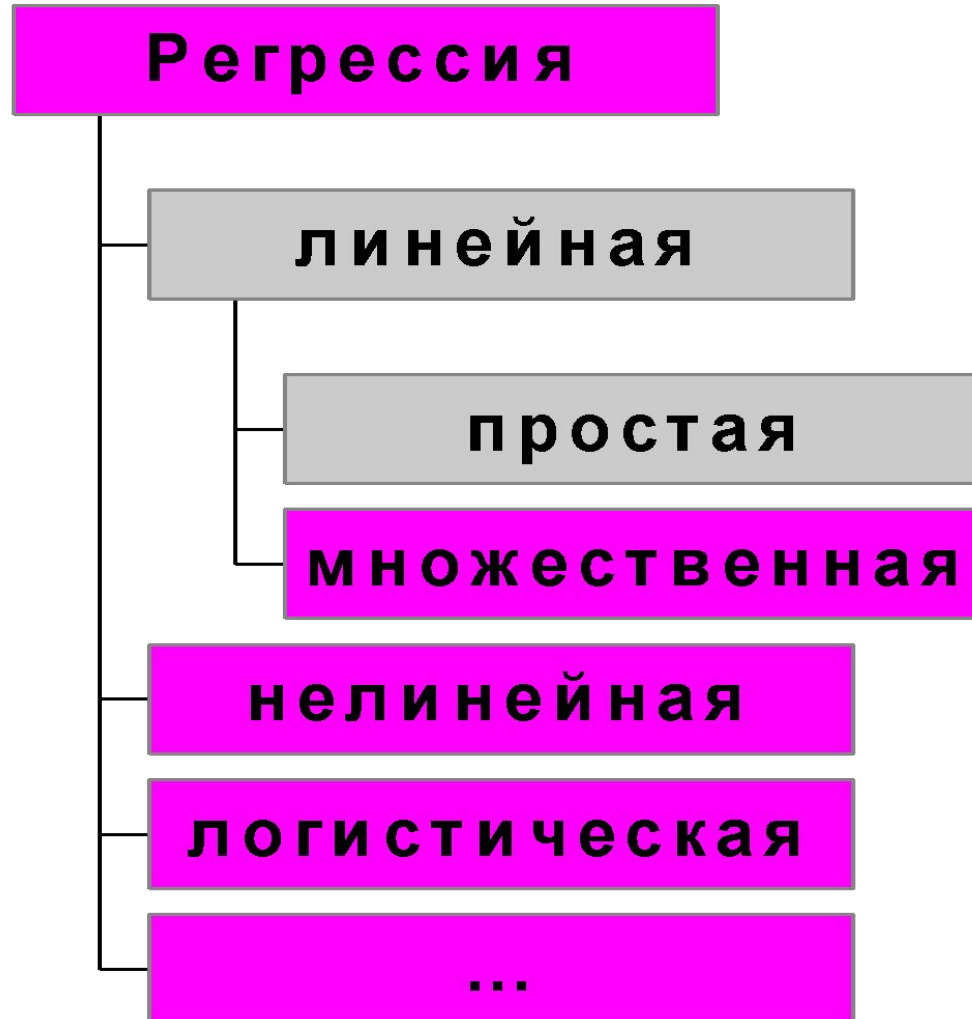


Регрессионный анализ





Рассмотрим сначала простую линейную регрессию.

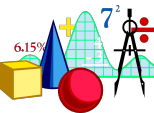




Ограничения

В случае простой линейной регрессии предполагается, что

- зависимая переменная одна и представлена по крайней мере в интервальной шкале**
- независимая переменная одна и представлена по крайней мере в интервальной шкале**

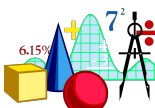




Пример 1: на диаграмме рассеяния показана зависимость показателя холестерина спустя 1 месяц после начала лечения (морковная диета) от исходного показателя.



Уравнение прямой помним еще из школы: Видно, что $y = bx + a$, множество точек, соответствующих b называется наблюдаемым регрессионным значением, коэффициентом концентрируется a b смещением вблизи прямой. В таком случае говорят о линейной связи. Задача состоит в нахождении a и b .

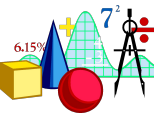




- Коэффициенты a и b вычисляются по формулам:

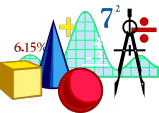
$$b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = r \frac{s_y}{s_x}, \quad a = \bar{y} - b\bar{x}$$

- Знак коэффициента регрессии совпадает со знаком коэффициента корреляции.



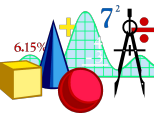


- **Равенство значения коэффициента регрессии нулю говорит об отсутствии линейной связи.**
- **Коэффициент регрессии показывает, насколько, в среднем, увеличится или уменьшится значение зависимой переменной y при увеличении независимой переменной x на 1.**



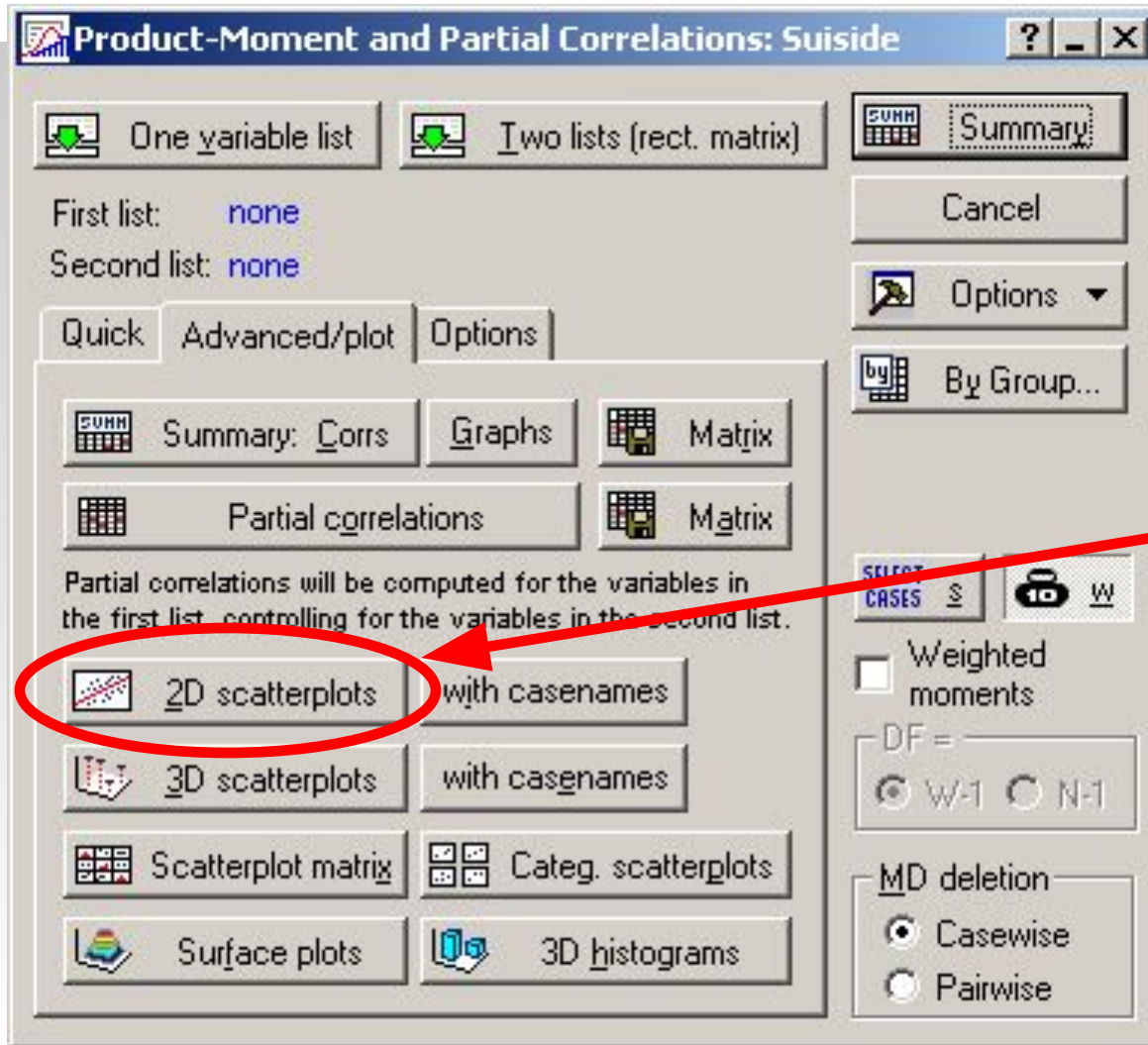


- Качество уравнения простой регрессии, его объясняющая способность измеряется **коэффициентом детерминации r^2** .
- Коэффициент детерминации показывает, какая доля дисперсии (изменчивости) переменной y объясняется влиянием независимой переменной x .

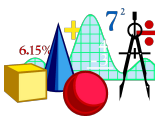




Уравнение простой линейной регрессии можно получить при построении диаграммы рассеяния:

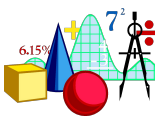


Надо только нажать эту кнопку





Уравнение регрессионной прямой и коэффициент корреляции





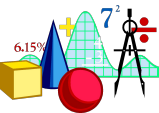
Для нашего примера 1

$$b=0,849; a=36,393$$



$$y=0,849x+36,393$$

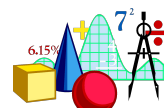
Теперь, зная, какой у вас уровень холестерина сейчас, можно предсказать, каков он будет через месяц лечения.





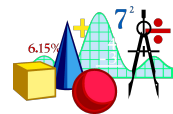
Доктор, у меня холестерин 310...

**Ничего страшного!
Через месяц морков-
ной диеты у Вас он
будет уже
 $0,849 \cdot 310 + 36,393 = 300!$**





Это было
просто!
Ерунда для
первого
курса!

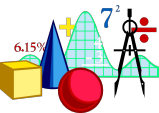




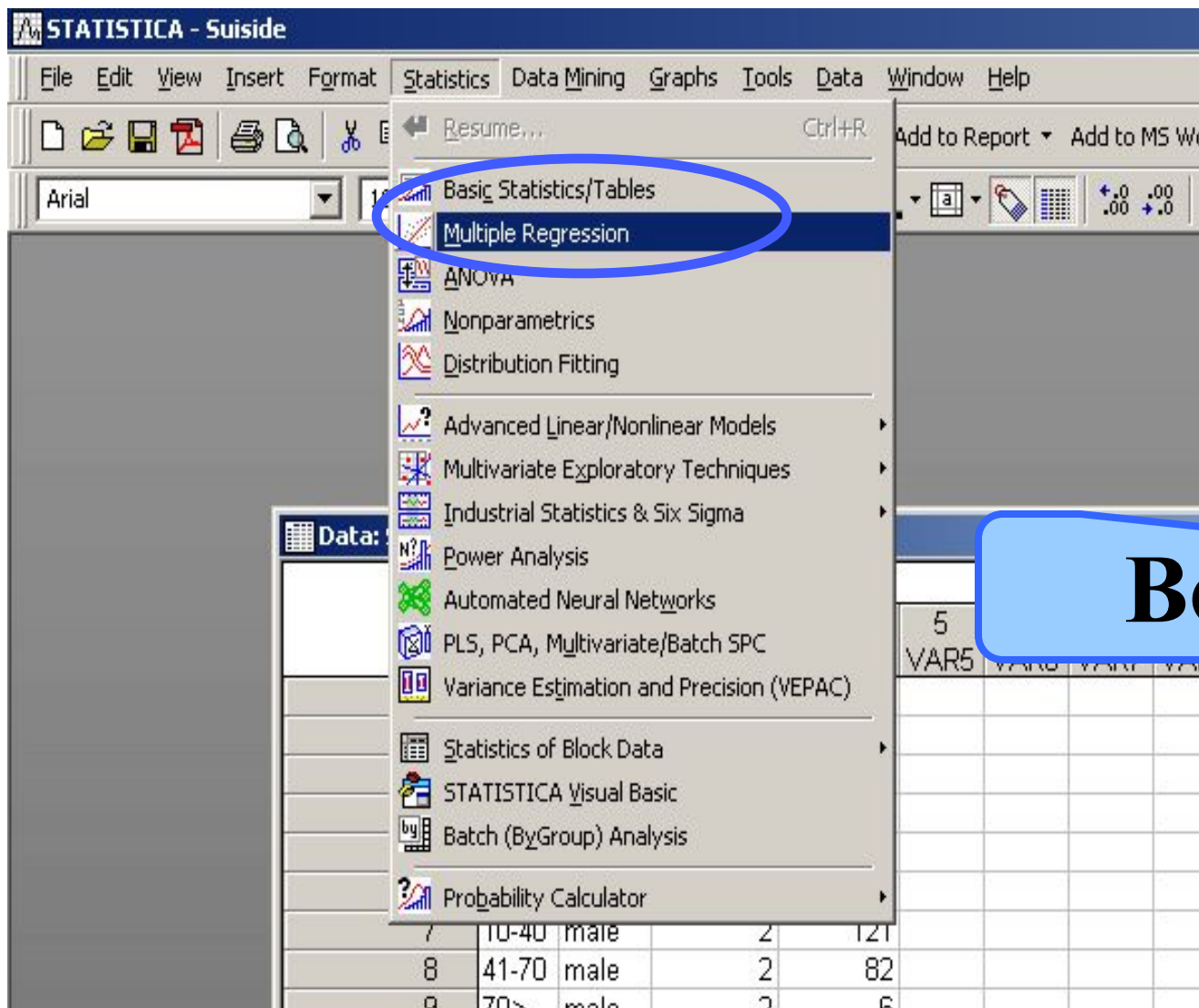
Модуль линейной регрессии

**Уравнение простой линейной регрессии
можно получить и в специальном модуле
программы STATISTICA.**

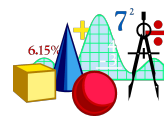
Он называется Multiple Regression



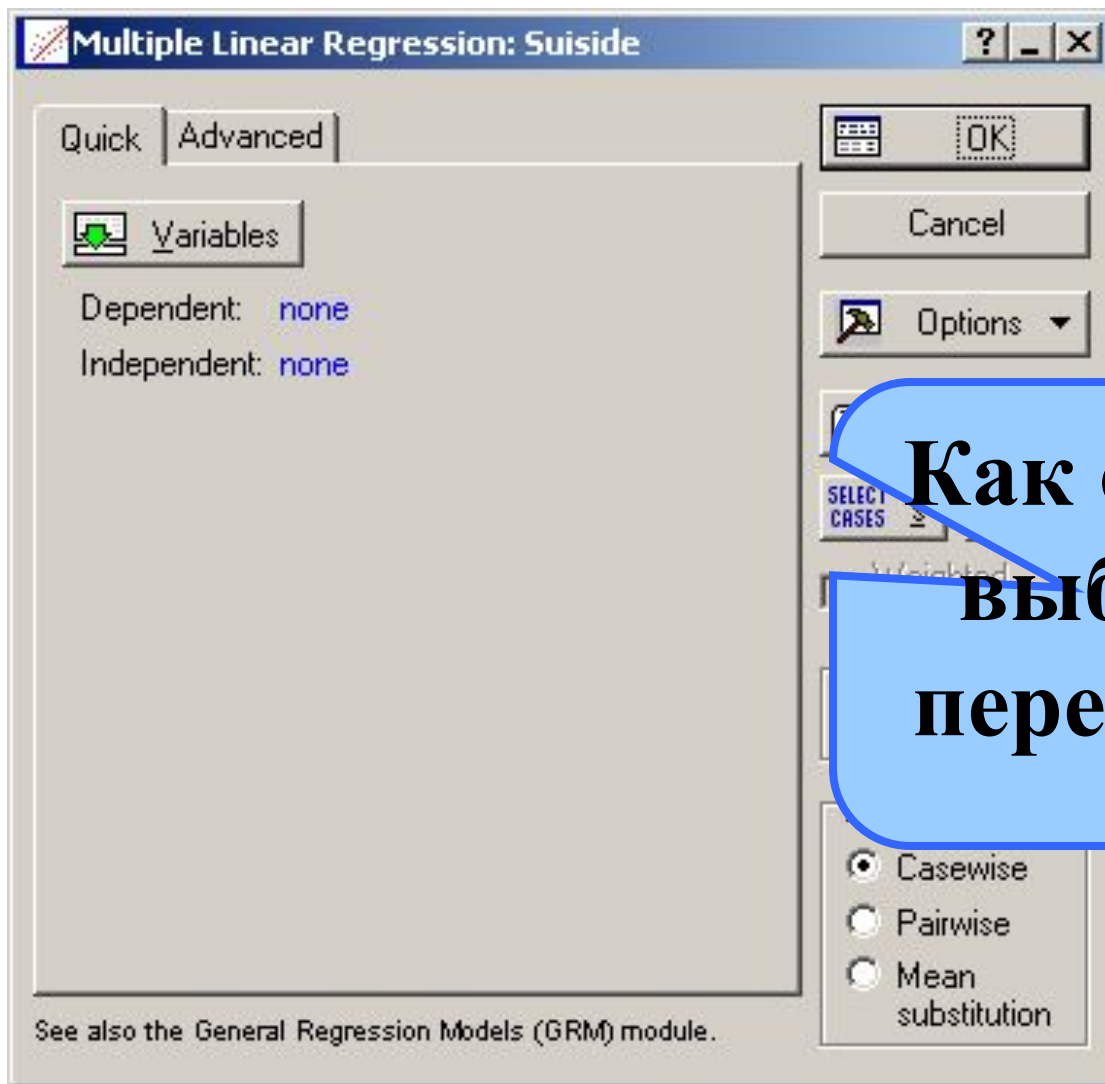
Модуль линейной регрессии



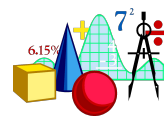
Вот он!



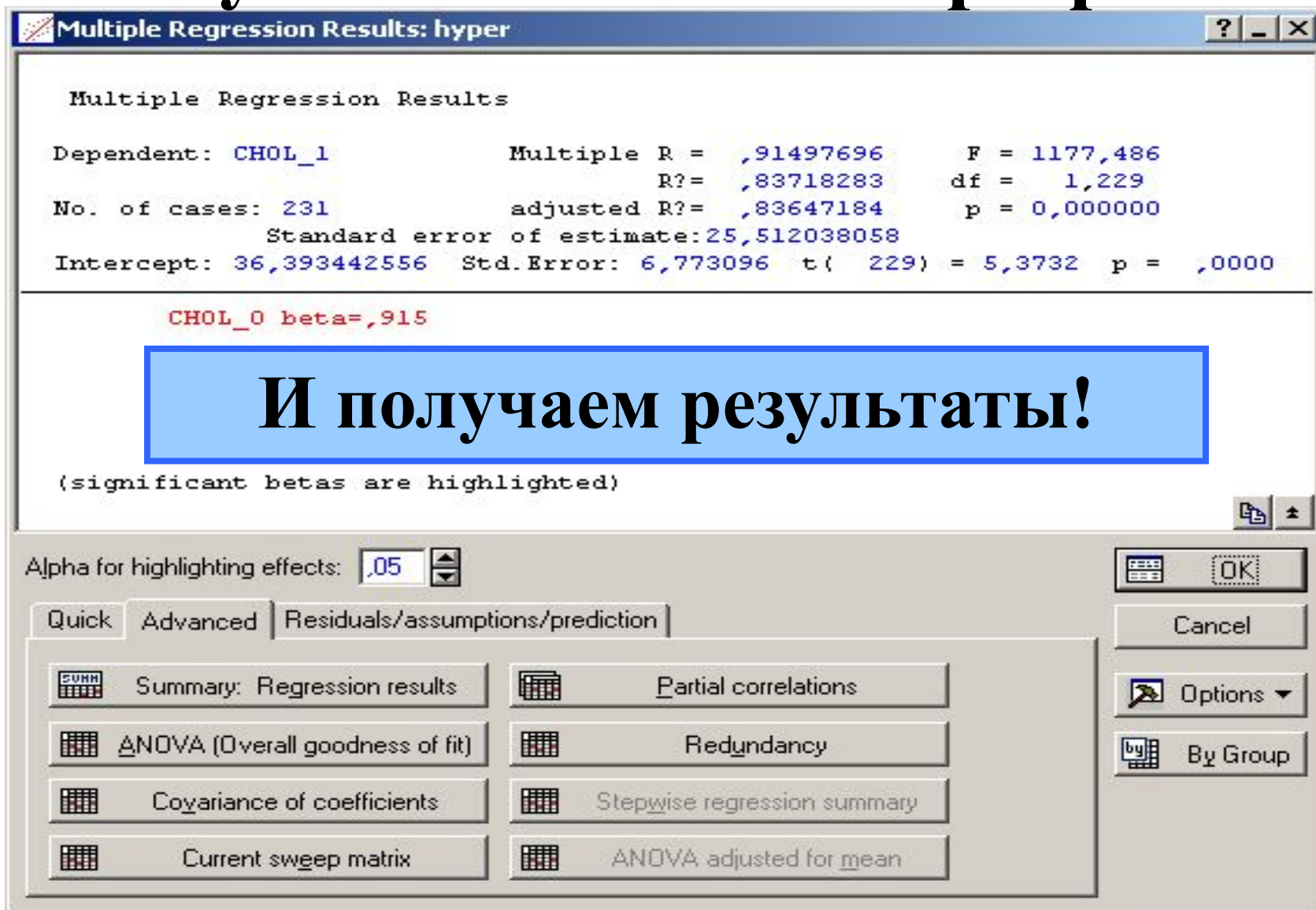
Модуль линейной регрессии



Как обычно,
выбираем
переменные



Результаты линейной регрессии



Multiple Regression Results: hyper

Multiple Regression Results

Dependent: CHOL_1 Multiple R = ,91497696 F = 1177,486
R² = ,83718283 df = 1,229
No. of cases: 231 adjusted R² = ,83647184 p = 0,000000
Standard error of estimate: 25,512038058

Intercept: 36,393442556 Std. Error: 6,773096 t(229) = 5,3732 p = ,0000

CHOL_0 beta = ,915

И получаем результаты!

(significant betas are highlighted)

Alpha for highlighting effects: .05

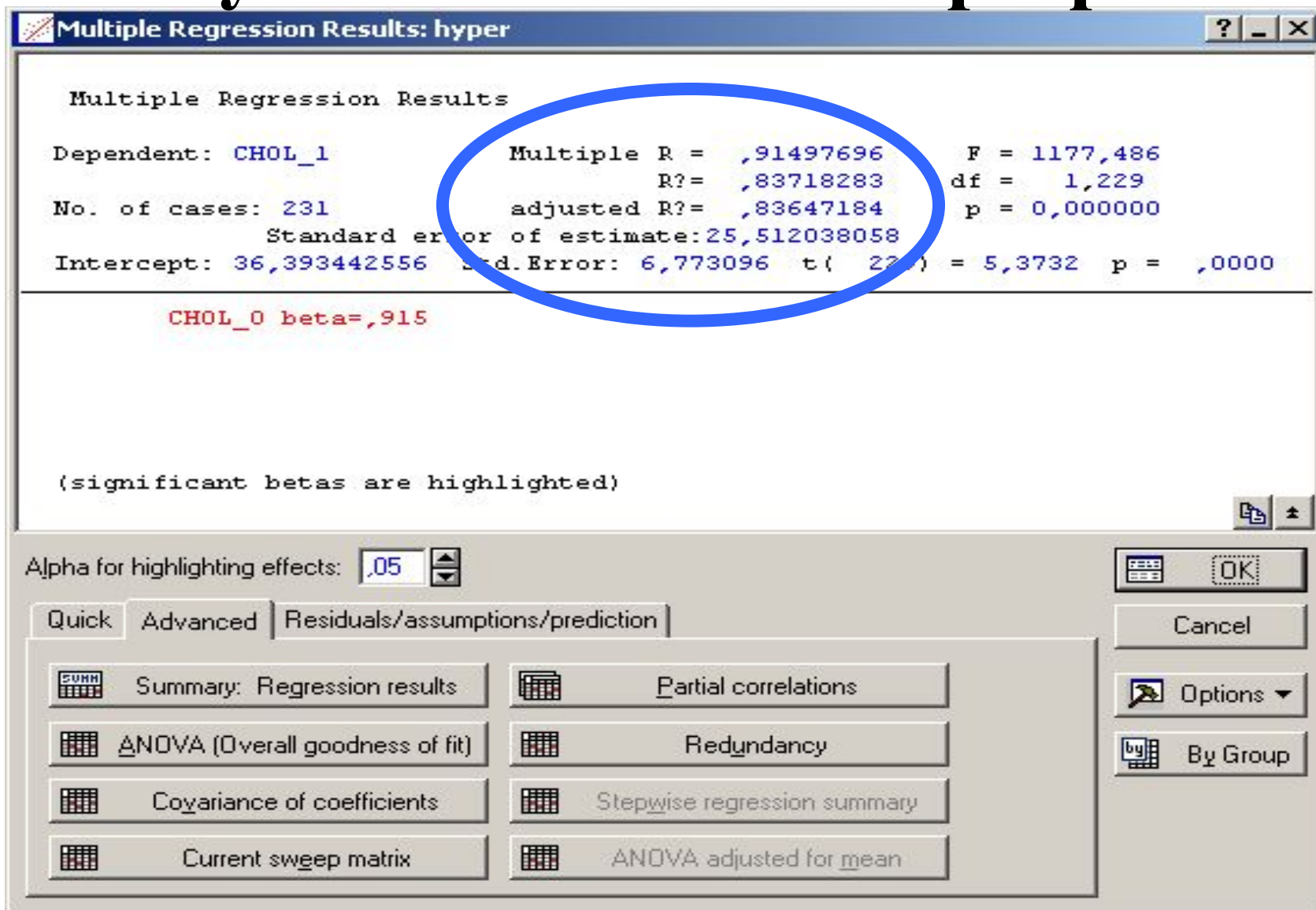
Quick | Advanced | Residuals/assumptions/prediction

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

OK | Cancel | Options | By Group



Результаты линейной регрессии

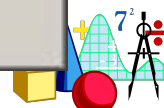


The screenshot shows the 'Multiple Regression Results: hyper' dialog box in SPSS. The main text area displays the following statistics:

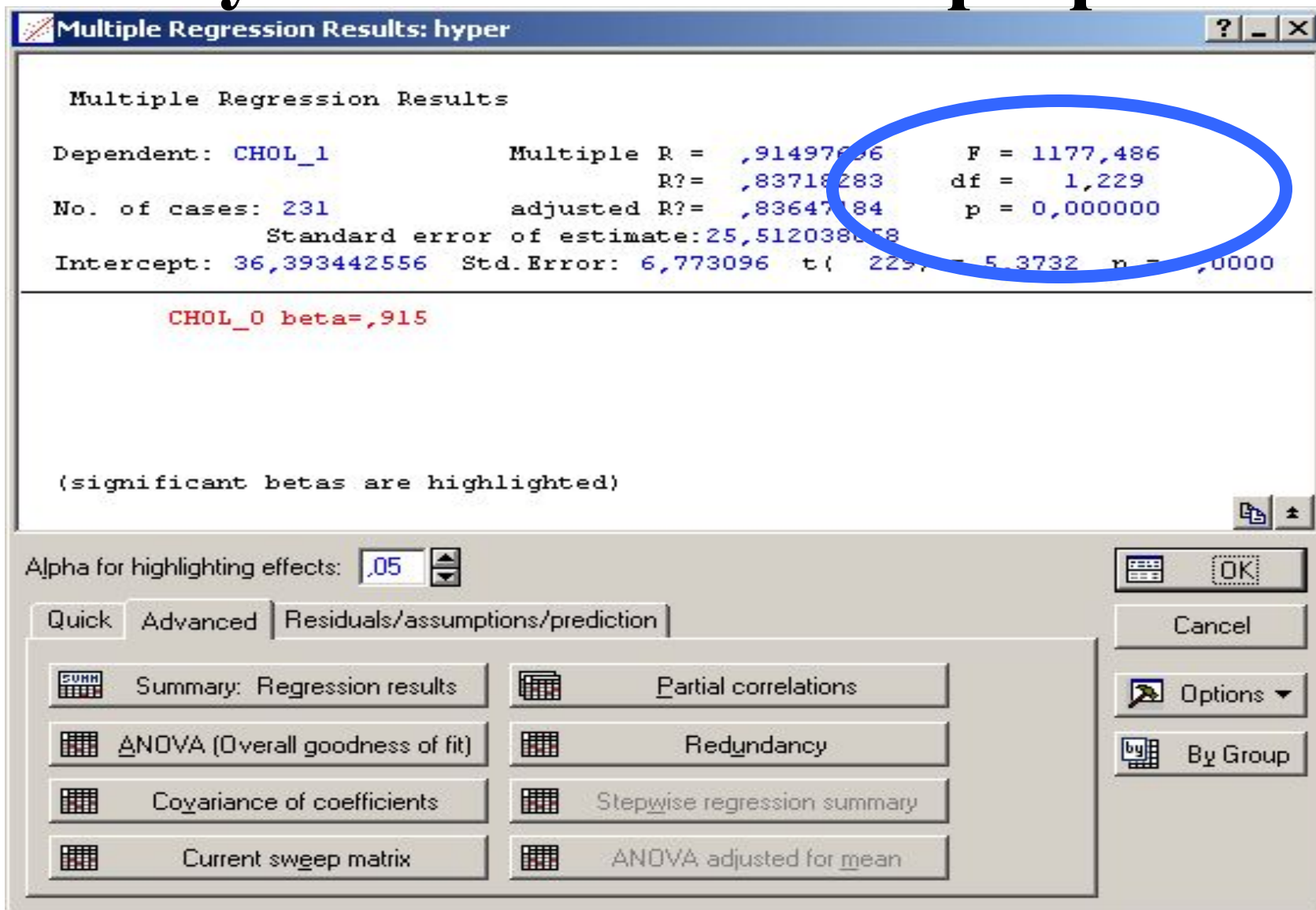
```
Multiple Regression Results
Dependent: CHOL_1
Multiple R = ,91497696    F = 1177,486
R^2 = ,83718283         df = 1,229
No. of cases: 231       adjusted R^2 = ,83647184    p = 0,000000
Standard error of estimate: 25,512038058
Intercept: 36,393442556 Std. Error: 6,773096 t( 22 ) = 5,3732 p = ,0000
```

A blue circle highlights the Multiple R, R-squared, and adjusted R-squared values. Below the main text, the text 'CHOL_0 beta=,915' is displayed in red. At the bottom, the 'Alpha for highlighting effects' is set to .05. The 'Quick' tab is selected, showing various options like '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'. The 'OK' button is highlighted.

(significant betas are highlighted)



Результаты линейной регрессии



The screenshot shows the 'Multiple Regression Results: hyper' dialog box in SPSS. The main text area displays the following regression statistics:

Dependent: CHOL_1	Multiple R = ,91497636	F = 1177,486
No. of cases: 231	R ² = ,83718283	df = 1,229
	adjusted R ² = ,83647184	p = 0,000000
	Standard error of estimate: 25,512038058	
Intercept: 36,393442556	Std. Error: 6,773096	t(229) = 5,3732 p = ,0000

Below the statistics, the text 'CHOL_0 beta=,915' is displayed in red, indicating a significant coefficient. A blue circle highlights the Multiple R, F, adjusted R², and p-value statistics.

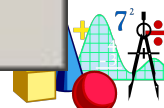
(significant betas are highlighted)

Alpha for highlighting effects: .05

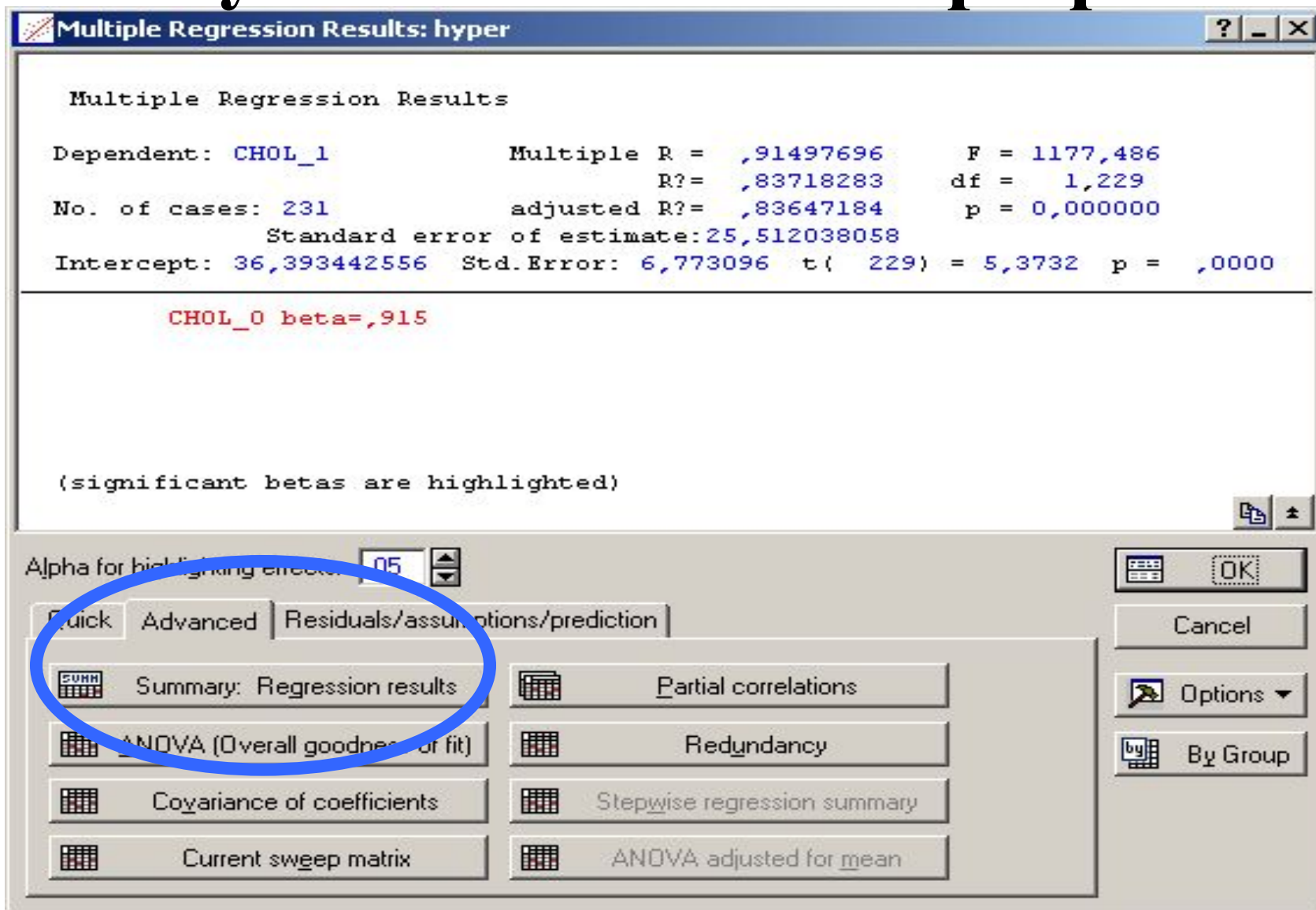
Buttons: Quick, Advanced, Residuals/assumptions/prediction

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

Buttons: OK, Cancel, Options, By Group



Результаты линейной регрессии



Multiple Regression Results: hyper

Multiple Regression Results

Dependent: CHOL_1 Multiple R = ,91497696 F = 1177,486
R² = ,83718283 df = 1,229
No. of cases: 231 adjusted R² = ,83647184 p = 0,000000
Standard error of estimate: 25,512038058

Intercept: 36,393442556 Std. Error: 6,773096 t(229) = 5,3732 p = ,0000

CHOL_0 beta = ,915

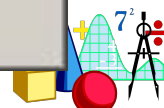
(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick | **Advanced** | Residuals/assumptions/prediction

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

OK | Cancel | Options | By Group

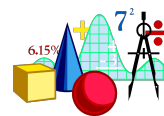


Результаты линейной регрессии

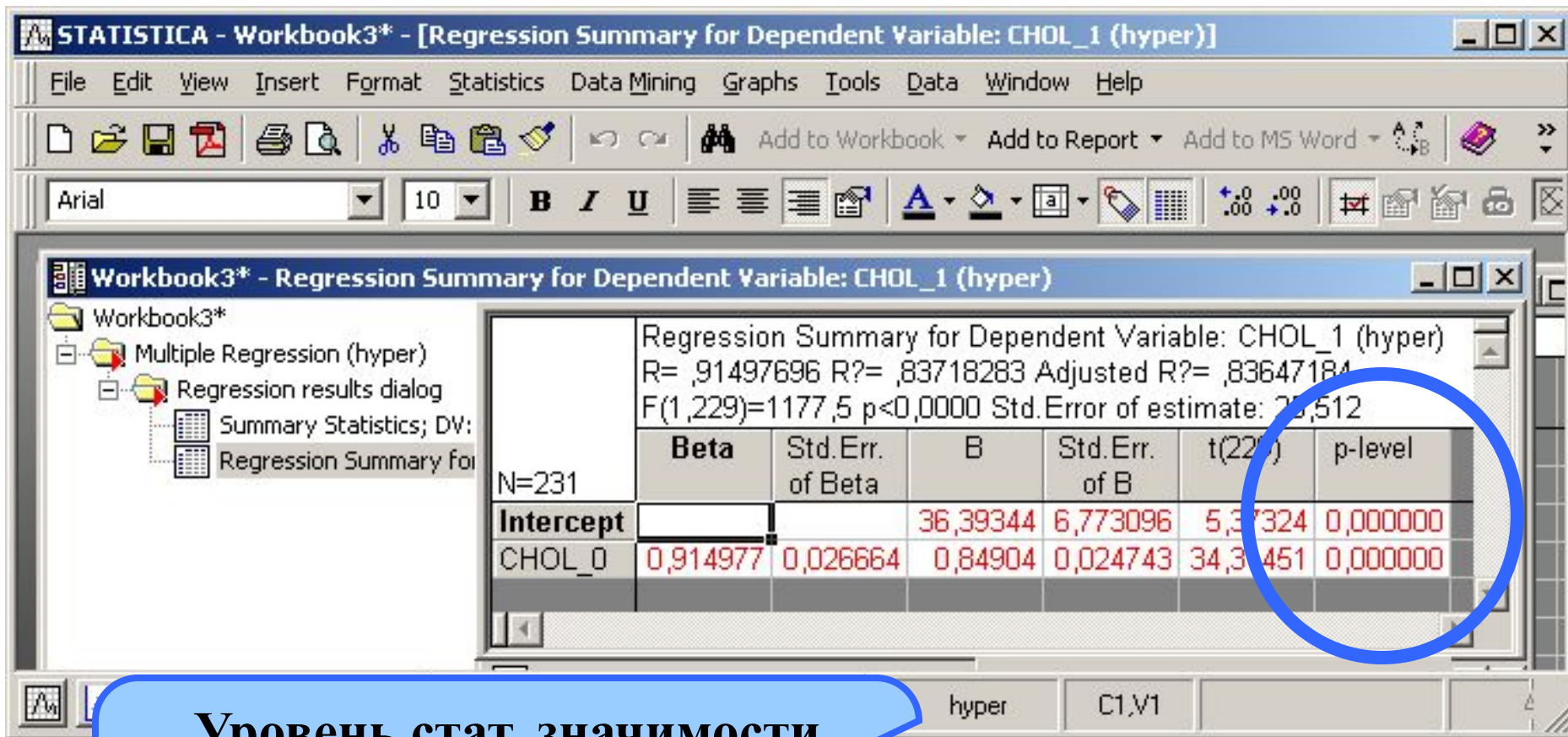
Regression Summary for Dependent Variable: CHOL_1 (hyper)
R= ,91497696 R²= ,83647184 Adjusted R²= ,83647184
F(1,229)=1177,5 < 0,0000 Std. Error of estimate: 25,512

	Beta	Std. Err. of Beta	B	Std. Err. of B	t(229)	p-level
N=231						
Intercept			36,39344	6,773096	5,37324	0,000000
CHOL_0	0,914977	0,026064	0,84904	0,024743	34,31451	0,000000

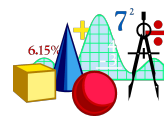
Коэффициенты
линейной
регрессии



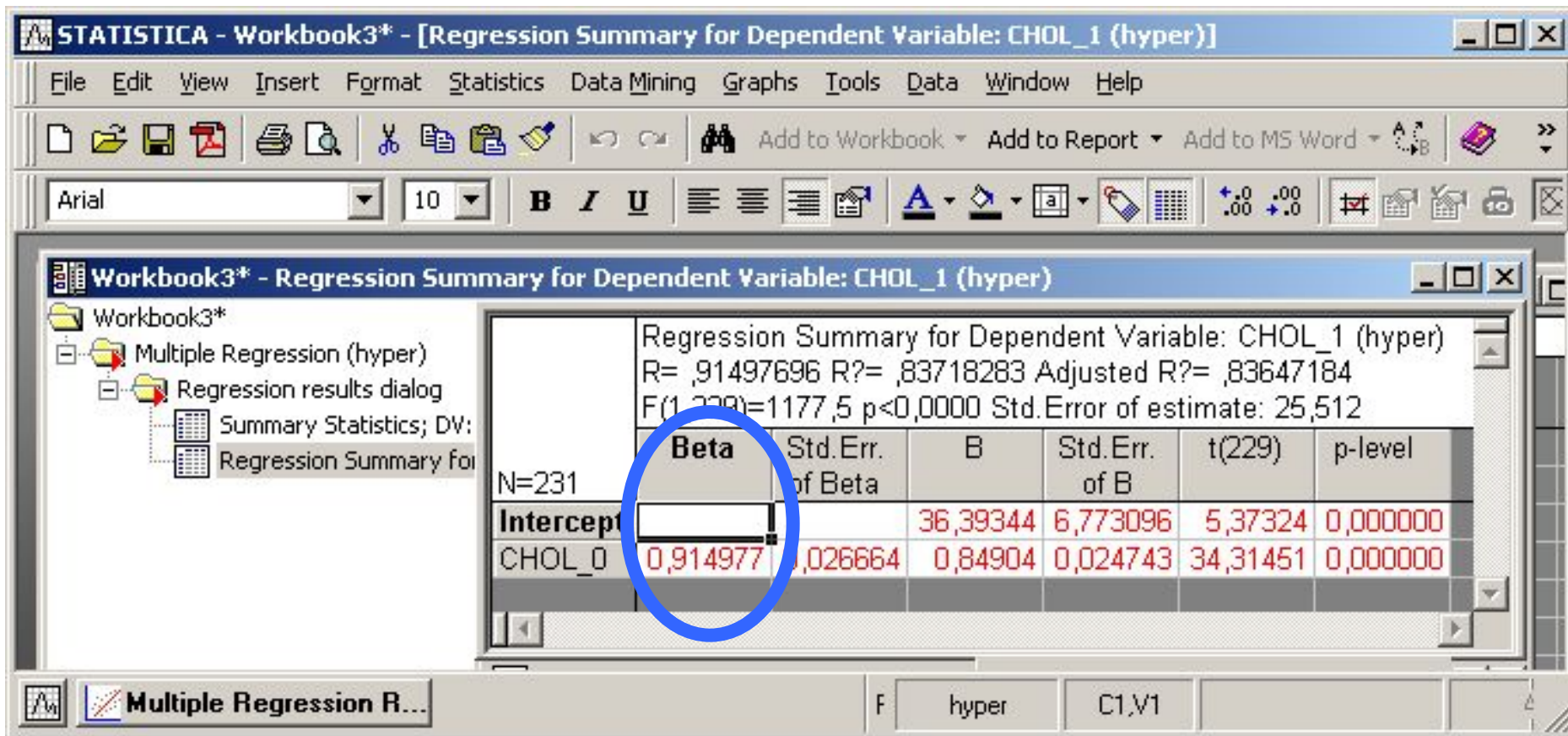
Результаты линейной регрессии



Уровень стат. значимости
коэффициентов линейной
регрессии



Результаты линейной регрессии



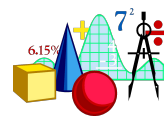
Коэффициенты β



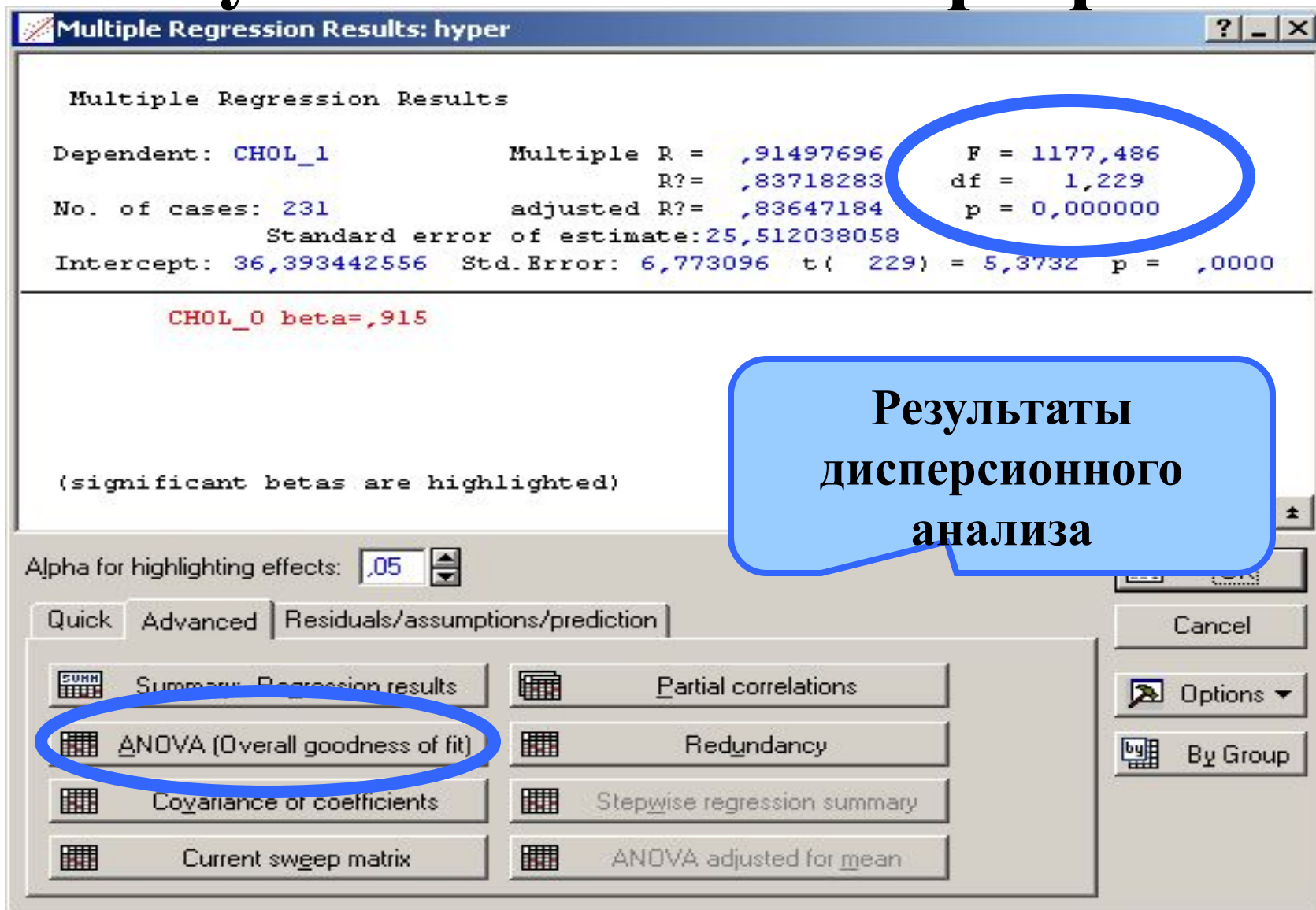
Результаты линейной регрессии

Коэффициенты β - это регрессионные коэффициенты, полученные в результате построения регрессионной модели в случае, когда все переменные предварительно нормированы (среднее=0, станд. отклон.=1)

ПРЕИМУЩЕСТВО: позволяют определить относительный вклад каждой независимой переменной в предсказании зависимой переменной.



Результаты линейной регрессии



Multiple Regression Results

Dependent: CHOL_1 Multiple R = ,91497696 F = 1177,486
R² = ,83718283 df = 1,229
No. of cases: 231 adjusted R² = ,83647184 p = 0,000000
Standard error of estimate: 25,512038058
Intercept: 36,393442556 Std. Error: 6,773096 t(229) = 5,3732 p = ,0000

CHOL_0 beta = ,915

(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

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

Cancel

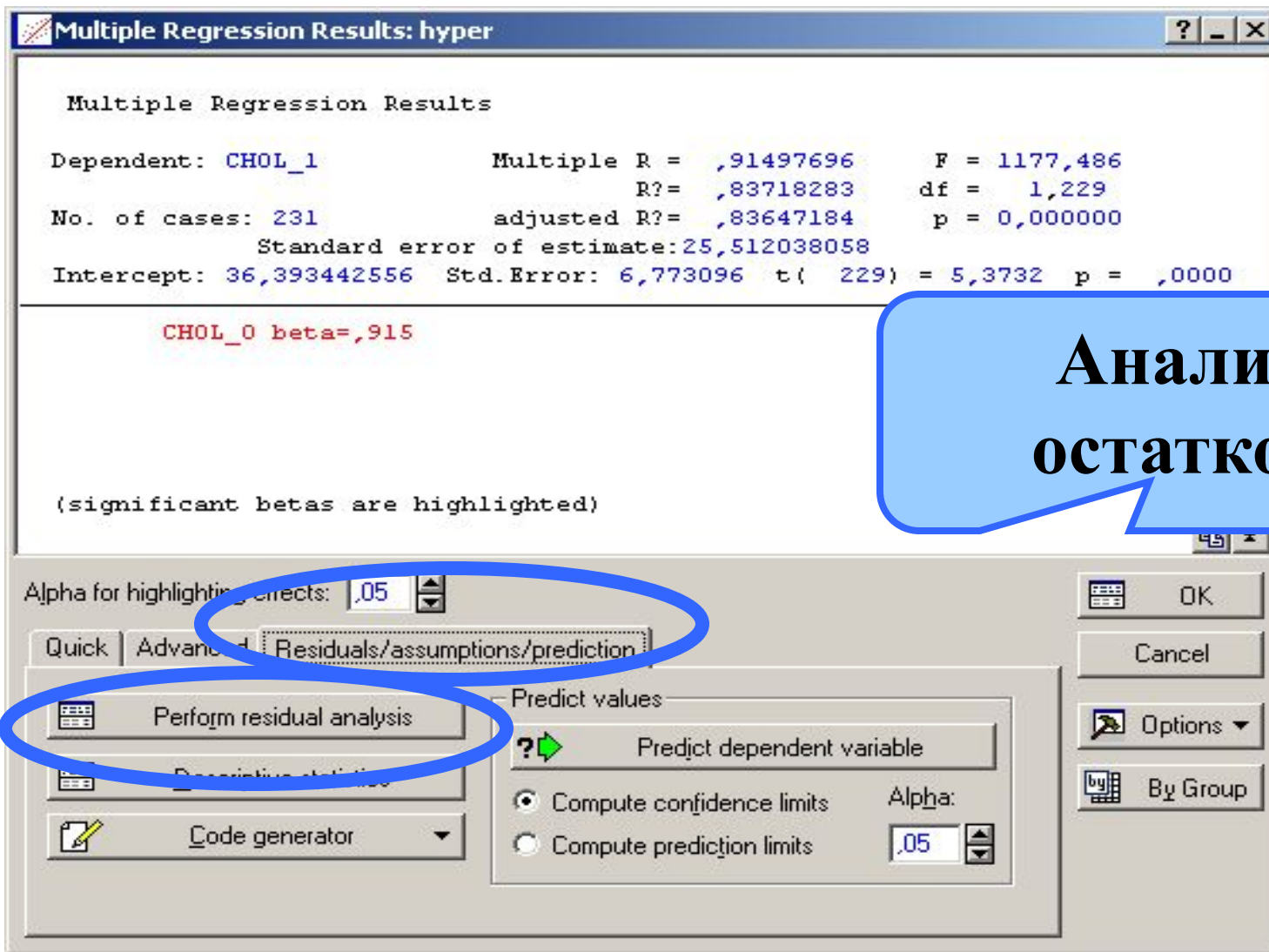
Options

By Group

Результаты дисперсионного анализа



Результаты линейной регрессии



The screenshot shows the 'Multiple Regression Results: hyper' dialog box in SPSS. The main window displays regression statistics for a dependent variable 'CHOL_1'. The statistics include Multiple R (0.91497696), F (1177.486), R-squared (0.83718283), adjusted R-squared (0.83647184), and a p-value of 0.000000. The standard error of estimate is 25.512038058. The intercept is 36.393442556 with a standard error of 6.773096 and a t-value of 5.3732 (p = 0.0000). A specific coefficient is highlighted: 'CHOL_0 beta = .915'. Below the statistics, it notes '(significant betas are highlighted)'. The dialog box has several tabs: 'Quick', 'Advanced', and 'Residuals/assumptions/prediction'. The 'Residuals/assumptions/prediction' tab is selected and contains options for 'Perform residual analysis', 'Descriptive statistics', and 'Code generator'. The 'Perform residual analysis' option is circled in blue. The 'Predict values' section is also visible, with 'Predict dependent variable' selected and 'Alpha' set to 0.05. The 'Alpha for highlighting effects' is also set to 0.05 and circled in blue. The 'OK' button is visible on the right side of the dialog box.

Multiple Regression Results

Dependent: CHOL_1 Multiple R = ,91497696 F = 1177,486
R² = ,83718283 df = 1,229
No. of cases: 231 adjusted R² = ,83647184 p = 0,000000
Standard error of estimate: 25,512038058
Intercept: 36,393442556 Std. Error: 6,773096 t(229) = 5,3732 p = ,0000

CHOL_0 beta = ,915

(significant betas are highlighted)

Alpha for highlighting effects: .05

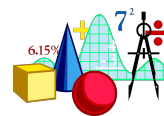
Quick | Advanced | **Residuals/assumptions/prediction**

Perform residual analysis
 Descriptive statistics
 Code generator

Predict values
Predict dependent variable
 Compute confidence limits Alpha: .05
 Compute prediction limits

OK
Cancel
Options
By Group

Анализ
остатков





Анализ остатков

Residual Analysis: hyper

Dependent: CHOL_1 Multiple R : ,91497696
R?: ,83718283
No. of cases: 231 adjusted R?: ,83647184
Standard error of estimate: 25,51203809
Intercept: 36,393442556 Std. Error: 6,773096 t

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

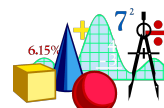
Histogram of residuals | Casewise plot of residuals | Residuals vs. independent var. | Histogram of observed

Type of residual

- Raw residuals
- Deleted residuals
- Standard residuals
- Cook's distances
- Mahalanobis distances

Cancel | Options | By Group

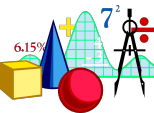
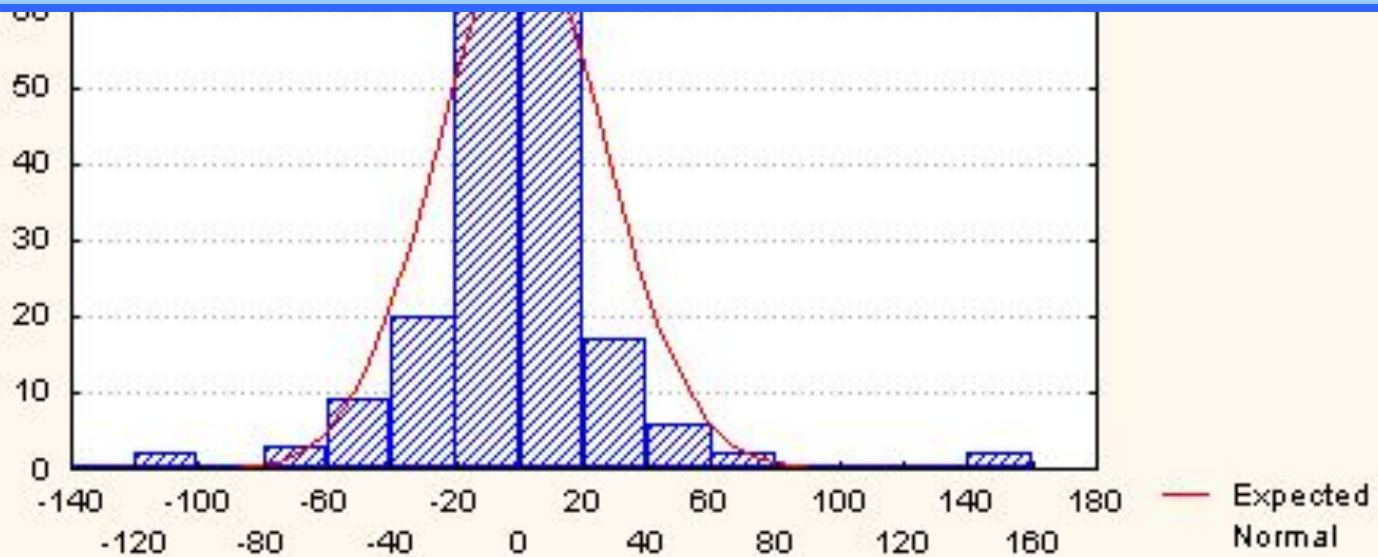
**Гистограмма
распределения
остатков**





Гистограмма распределения остатков

Распределение должно быть нормальным





Анализ остатков

График предсказанных и наблюдаемых (эмпирических) значений

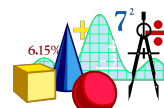
Residual Analysis: hyper

Dependent: CHOL_1 Multiple R : ,91497696 F = 1177,486
No. of cases R?: ,83718283 df = 1,229

Quick | Advanced | Statistics | Diagnostics

Predicted vs. residuals	Observed vs. squared residuals
Predicted vs. squared residuals	Residuals vs. deleted residuals
Predicted vs. observed	Bivariate correlation
Observed vs. residuals	Partial residual plot

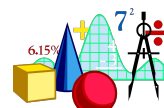
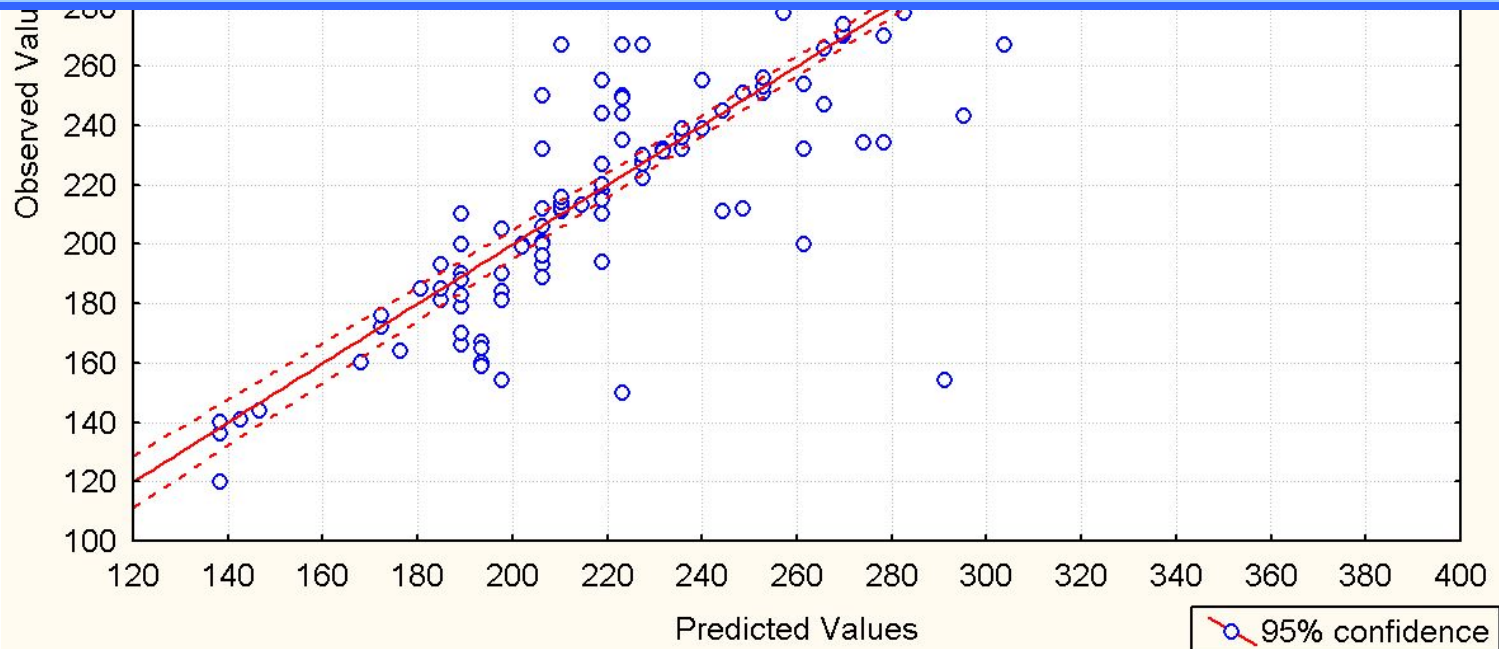
Cancel
Options
By Group





Анализ остатков

Эти значения должны лежать вдоль одной прямой





Анализ остатков

График вероятностей нормального распределения?

Residual Analysis: hyper

Dependent: CHOL_1 Multiple R : ,91
R?: ,83
No. of cases: 231 adjusted R?: ,83
Standard error of estimate: 25,5
Intercept: 36,393442556 Std. Error: 6,77

Quick | Advanced | **Residuals** | Predicted | Scatterplots | Histograms | Outliers

Normal plot of residuals

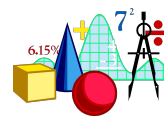
Half-normal plot

Detrended normal plot

Cancel

Options ▾

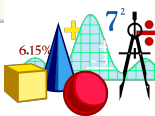
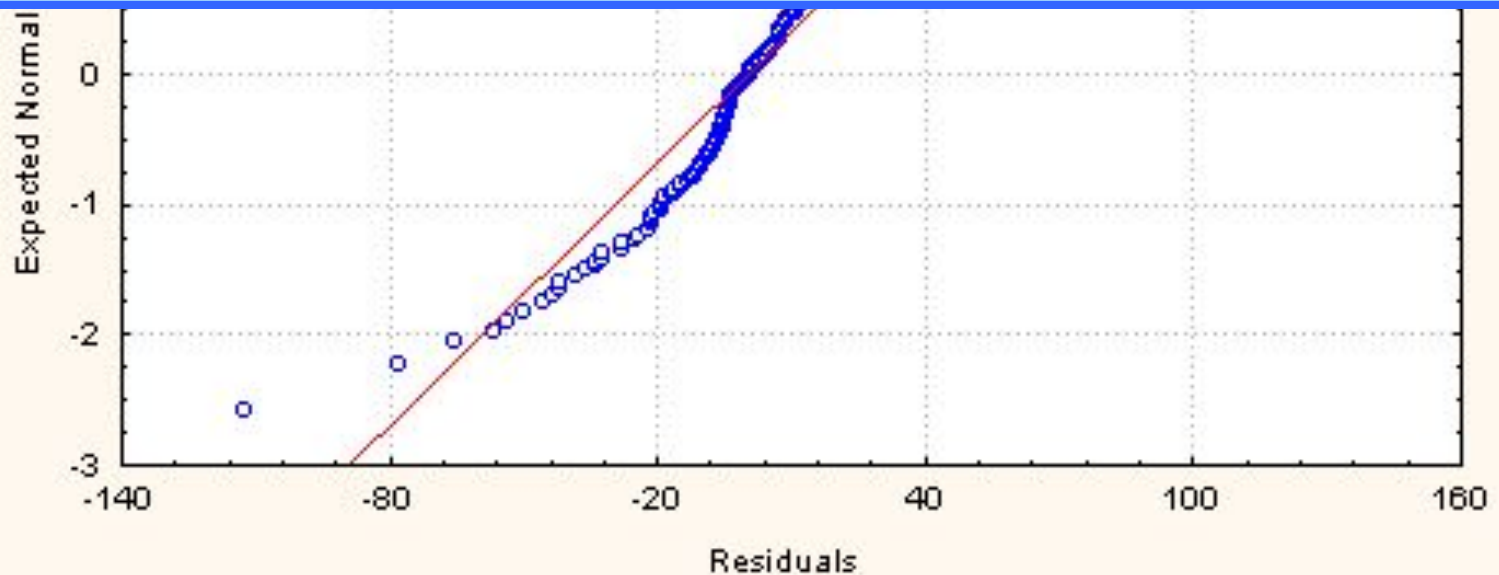
By Group



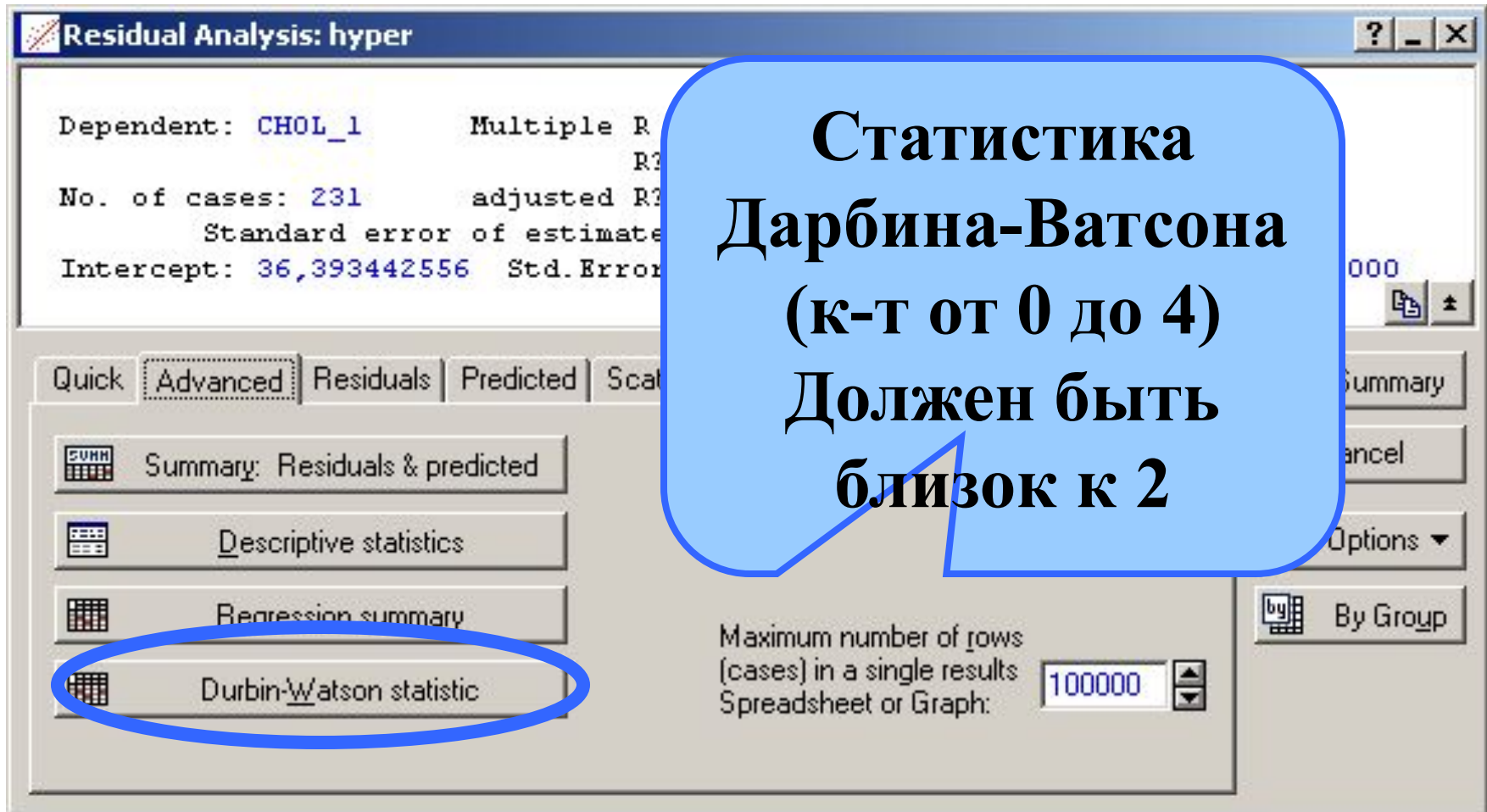


Анализ остатков

Эти значения должны лежать вдоль одной прямой



Анализ остатков



Residual Analysis: hyper

Dependent: CHOL_1 Multiple R

No. of cases: 231 adjusted R²

Intercept: 36,393442556 Std. Error

Quick **Advanced** Residuals Predicted Scatter

Summary: Residuals & predicted

Descriptive statistics

Regression summary

Durbin-Watson statistic

Maximum number of rows (cases) in a single results Spreadsheet or Graph: 100000

Summary

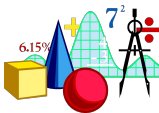
Cancel

Options

By Group

000

Статистика
Дарбина-Ватсона
(к-т от 0 до 4)
Должен быть
близок к 2



Результаты линейной регрессии

Multiple Regression Results: hyper

Multiple Regression Results

Dependent: CHOL_1 Multiple R =
R² =

No. of cases: 231 adjusted R² =

Standard error of estimate: 25

Intercept: 36,393442556 Std. Error: 6,7730

CHOL_0 beta=,915

(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick | Advanced | **Residuals/assumptions/prediction**

Perform residual analysis

Descriptive statistics

Code generator

Predict values

Predict dependent variable

Compute confidence limits Alpha: .05

Compute prediction limits

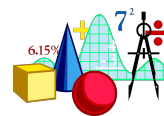
OK

Cancel

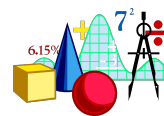
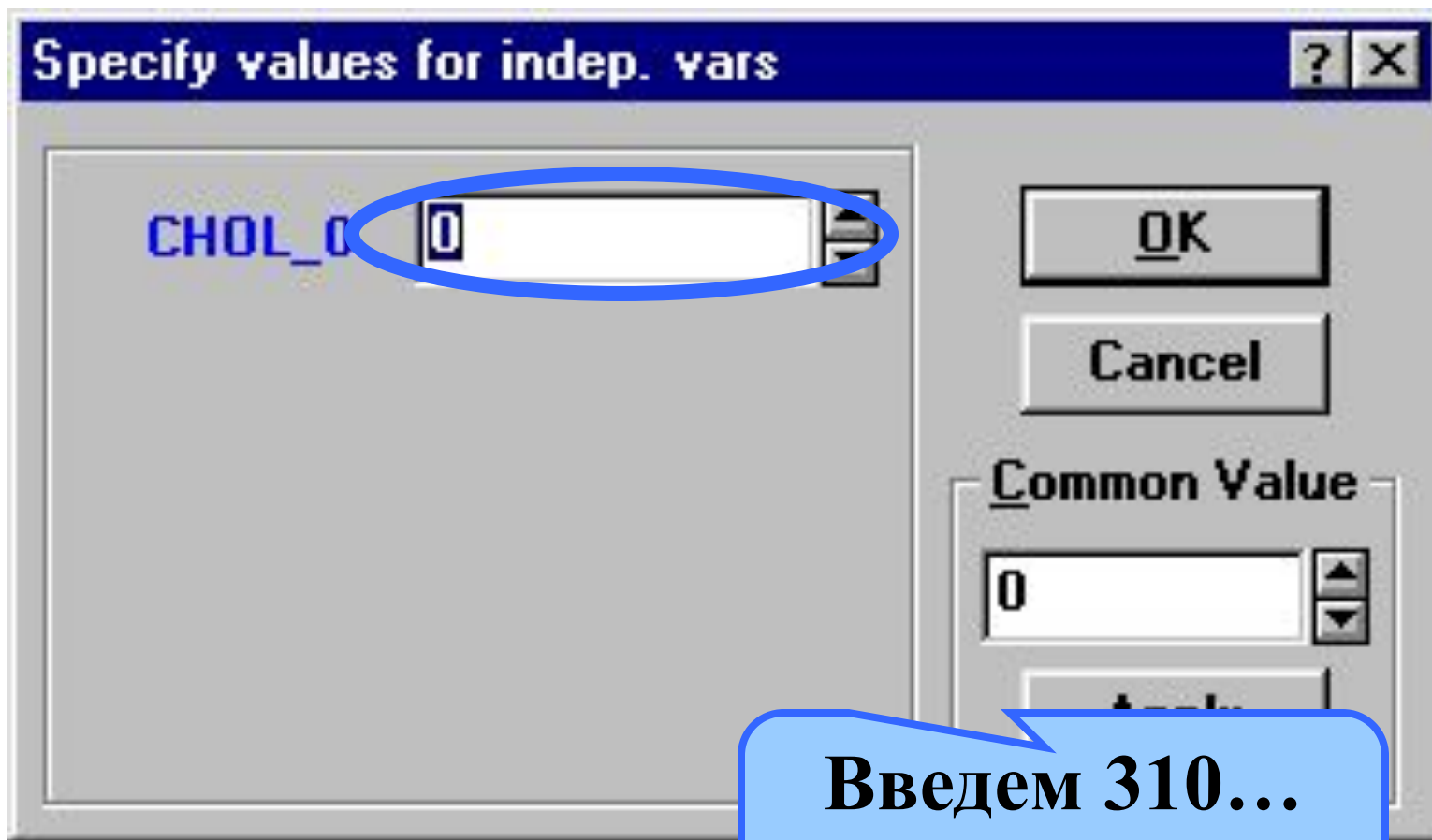
Options

By Group

Прогноз



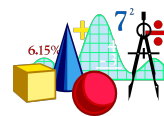
Результаты линейной регрессии



Результаты линейной регрессии

и получим 300 через месяц
морковной диеты + 95%
дов. интервал

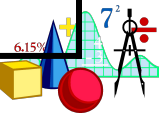
variable	B-Weight	Value	B-Weight * Value
CHOL_0	,849041	310,0000	263,2028
Intercpt			36,3934
Predictd			299,5962
-95,0%CL			295,6327
+95,0%CL			303,5598





Пример 2

Возраст (лет)	Словарный запас (число слов)
1.0	3
1.5	22
2.0	272
2.5	446
3.0	896
3.5	1222
4.0	1540
4.5	1870
5.0	2072
6.0	2562





Пример 2

Regression Summary for Dependent Variable: VOCAB (new.sta)

MULTIPLE R= ,99260871 **RI= ,98527205** Adjusted RI= ,98343105
REGRESS. F(1,8)=535,18 p<,00000 Std.Error of estimate: 116,74

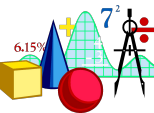
N=10	BETA	St. Err. of BETA	B	St. Err. of B	t(8)	p-level
Intercpt			-763,857	88,24996	-8,65561	,000025
AGE	,992609	,042907	561,926	24,29000	23,13406	,000000

Словарный запас = 562*возраст – 764

В 7 лет - 3170 слов

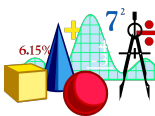
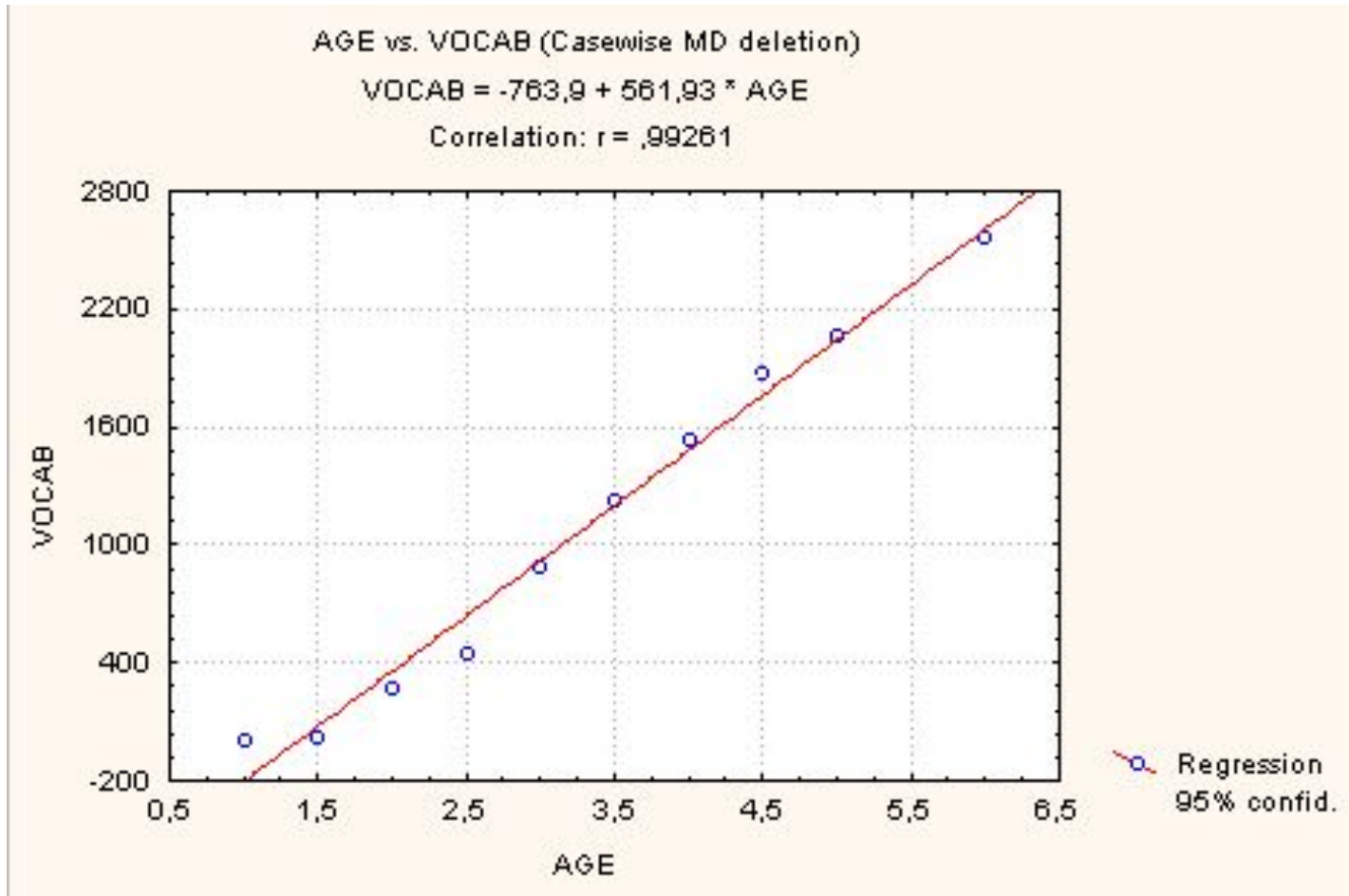
В 10 лет – 4855 слов

...





Пример 2



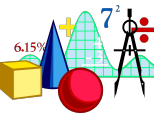


Пример 2

А что было, когда ребенок только родился?

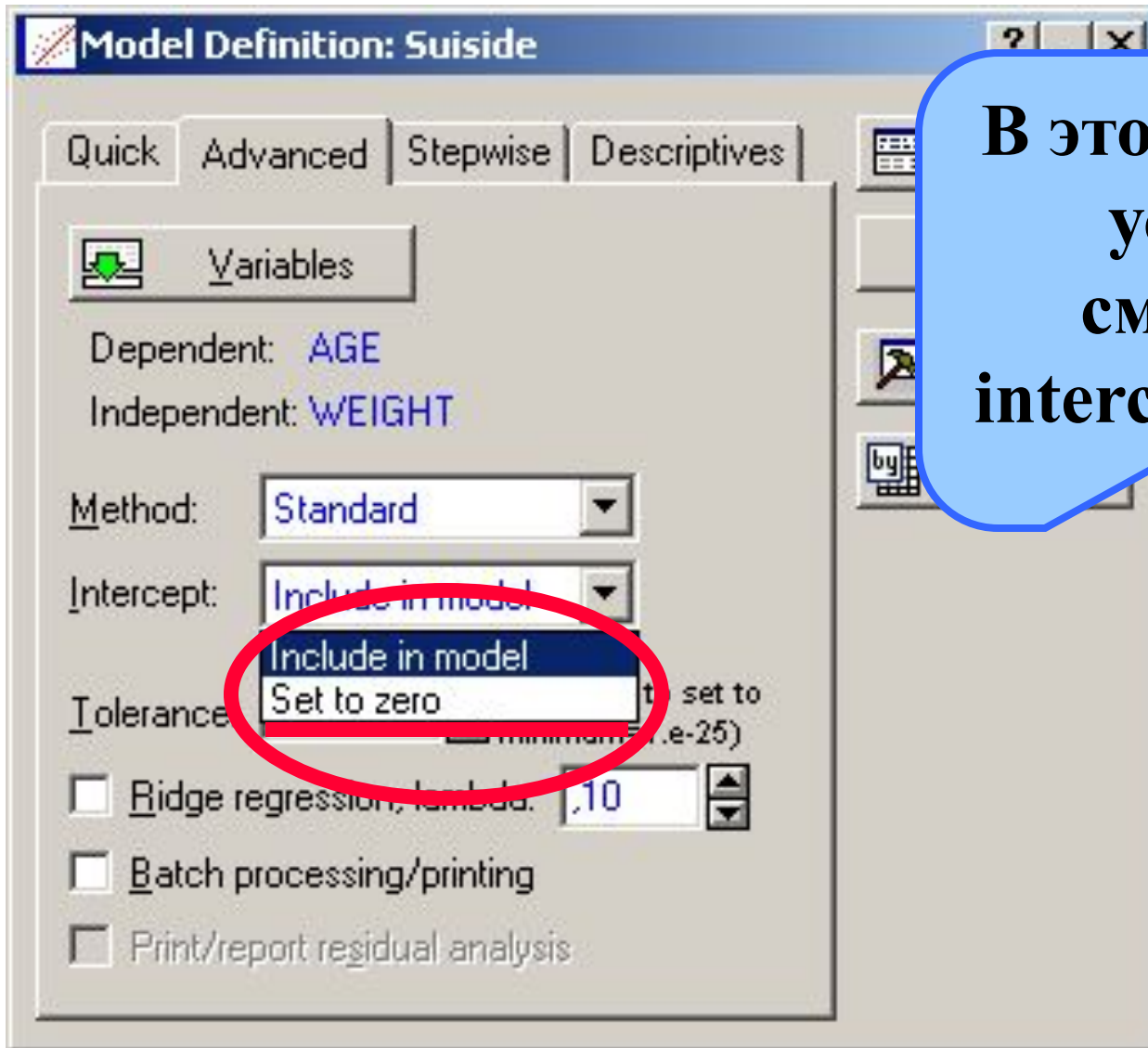
**В 0 лет словарный запас =
= $562 * \text{возраст} - 764 = -764$ слова!**

**Поэтому есть возможность установить
смещение = 0**

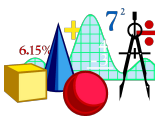




Пример 2

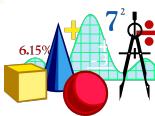
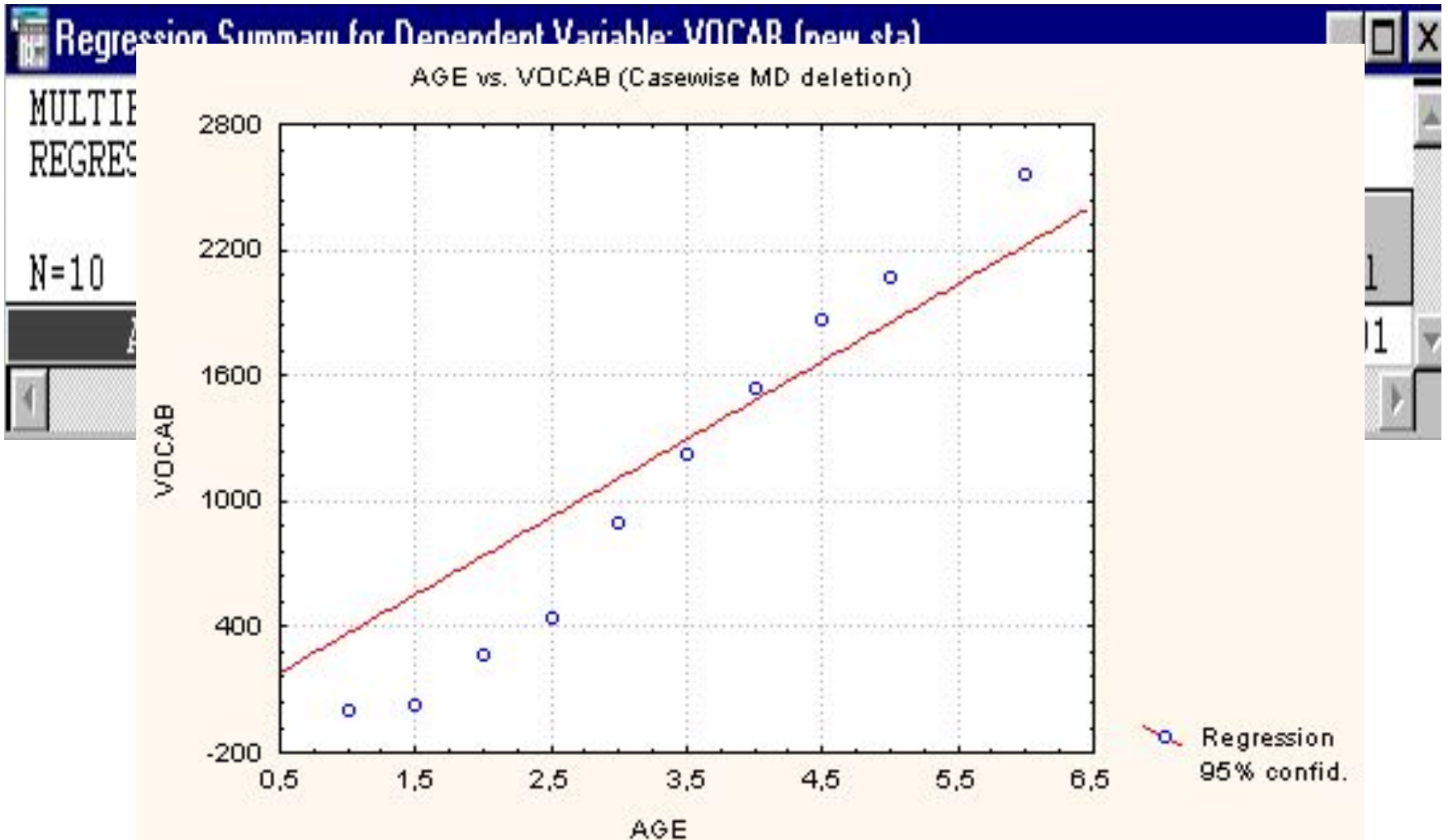


**В ЭТОМ ОКНЕ МОЖНО
УСТАНОВИТЬ
СМЕЩЕНИЕ=0:
intercept: set to zero**



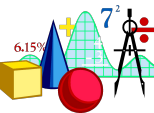


Пример 2



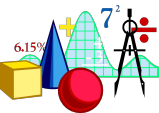
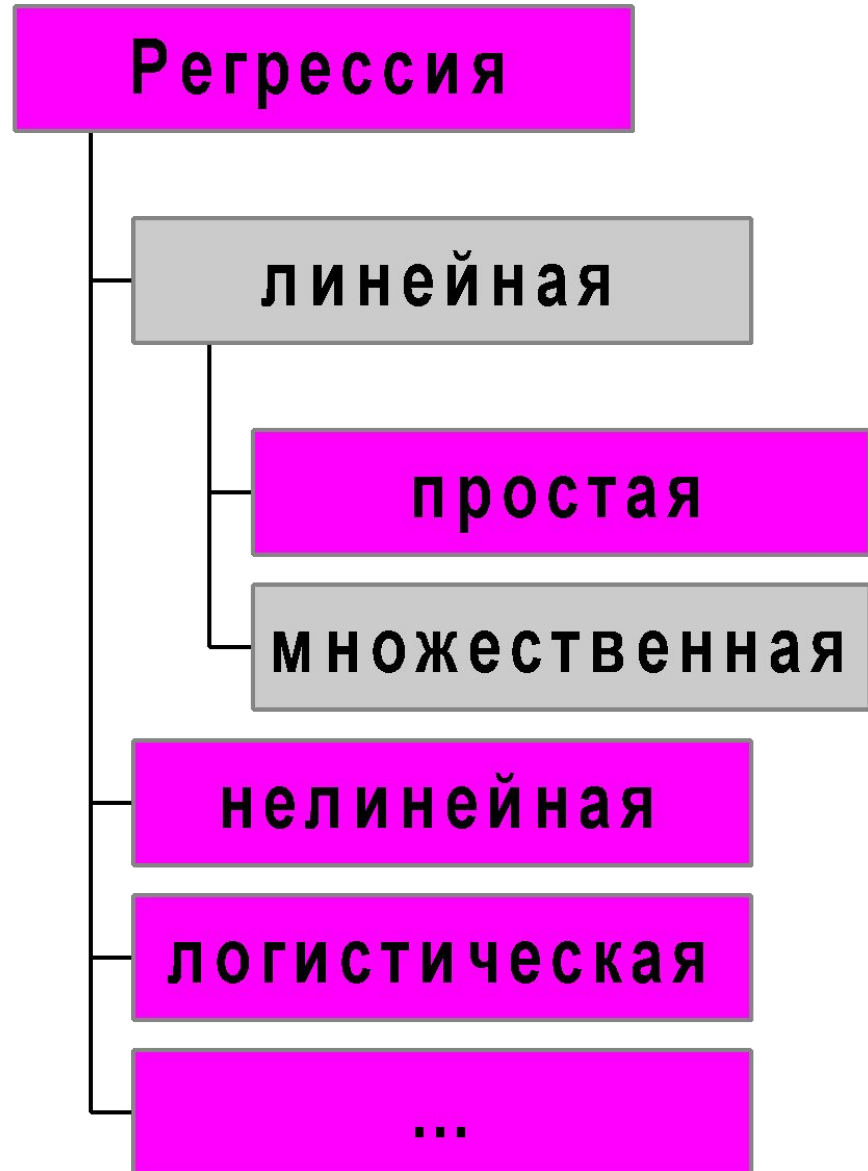


Фух!
Достаточно
про простую
линейную
регрессию!





Бывает, что действие
зависимой
переменной не
может быть
объяснено только
одной причиной
(независимой)
переменной. Тогда
воспользуемся
услугами
множественной
регрессии:





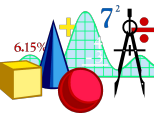
Уравнение множественной регрессии очень похоже на уравнение простой линейной регрессии:

$$Y = b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_n x_n + a$$

b_i - регрессионные коэффициенты

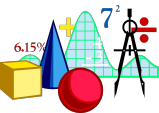
X_i - независимые переменные, их столько, сколько вам не лень придумать или измерить

a - свободный член





**Наша задача заключается в
определении
коэффициентов b_i и a**

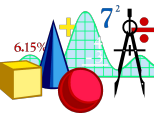




Ограничения

В случае множественной линейной регрессии предполагается, что

- зависимая переменная одна и представлена по крайней мере в интервальной шкале**
- независимых переменных несколько и они представлены либо в интервальной шкале, либо в шкале равных отношений, либо в шкале наименований (!)**

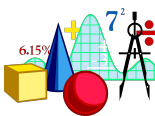




Это тоже можно сделать в модуле Multiple Regression

The screenshot shows the STATISTICA software interface. The 'Statistics' menu is open, and 'Multiple Regression' is highlighted. A red arrow points to this menu item. The interface includes a menu bar (File, Edit, View, Insert, Format, Statistics, Data Mining, Graphs, Tools, Data, Window, Help), a toolbar, and a data table at the bottom.

	5	6	7	8
	VAR5	VAR6	VAR7	VAR8
7	10-40	male	2	121
8	41-70	male	2	82
9	70s	male	2	6



Пример № 3

(использование множественной
регрессии):

анализ данных по недвижимости

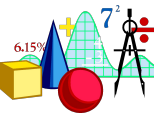
**Рассматривались данные
по двухкомнатным квартирам
Число квартир в базе - 6286**





Информация по каждой квартире:

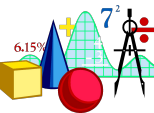
- Цена квартиры (в тыс. \$),
- Общая площадь (в м²),
- Жилая площадь (в м²),
- Площадь кухни (в м²),
- Расстояние от центра (в км),
- Способ добраться до метро
(бинарная переменная, принимающая значение 1- пешком, 0- на транспорте).





Информация по каждой квартире:

- **Тип постройки здания**
(бинарная переменная:
1- кирпичный дом, 0- панельный дом)
- **Высота расположения квартиры**
(1 - если квартира находится
не на 1 или последнем этаже,
0 - в противном случае).



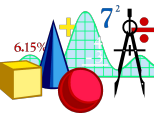


Переменные регрессионного анализа

В приведенной базе данных есть дихотомические(есть-нету) (бинарные) переменные. *Это переменные, принимающие всего два значения.*

Дихотомические переменные ведут себя так же, как интервальные!!!(ср.арифметическое и дисперсия).

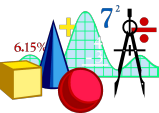
Для них среднее арифметическое имеет смысл и можно считать к-т корреляции Пирсона!





Задачи исследования

- Провести анализ влияния характеристик квартиры на ее цену
- Построить модель зависимости стоимости квартиры от исследуемых параметров и численно оценить коэффициенты модели a и b





Начинаем анализ

Выбор переменных

Выбор метода

Multiple Linear Regression: FLAT

Quick | Advanced

Variables

Dependent: PRICE
Independent: TOTSP-BAL

Input file: Raw Data

Advanced options (stepwise or ridge regression)

Review descriptive statistics, correlation matrix

Extended precision computations

Batch processing/reporting

Print/report residual analysis

Specify all variables for the analysis; additional models (indep./dep. vars) can be specified later. For stepwise regression etc. check the advanced options check box.

See also the General Regression Models (GRM) module.

OK

Cancel

Options

Open Data

SELECT CASES

Weighted moments

DF =

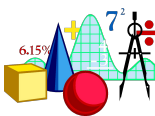
W-1 N-1

MD deletion

Casewise

Pairwise

Mean substitution





Начинаем анализ

Выбор переменных

Выбор метода

Model Definition: FLAT

Quick | Advanced | Stepwise | Descriptives

Variables

Dependent: PRICE

Independent: TOTSP-BAL

Method: Standard

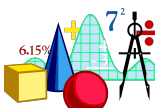
- Standard
- Forward stepwise
- Backward stepwise

OK

Cancel

Options

By Group





Начинаем анализ

Выбор переменных

Model Definition: FLAT

Quick | Advanced | Stepwise | Descriptives

Variables

Dependent: PRICE
Independent: TOTSP-BAL

Method: Standard

Intercept: Include in model

Tolerance: Set to zero (to set to minimum=1.e-25)

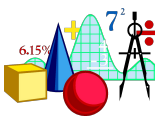
Ridge regression; lambda: .10

Batch processing/printing

Print/report residual analysis

OK
Cancel
Options
By Group

Пересечение с осью У





Начинаем анализ

Select dependent and independent variable lists: ? X

1-PRICE 3-PODSP 4-LIVSP 5-KITSP 6-DIST 7-WALK 8-BRICK 9-FLOOR 10-METRDIST 11-TEL 12-BAL	1-PRICE 3-PODSP 4-LIVSP 5-KITSP 6-DIST 7-WALK 8-BRICK 9-FLOOR 10-METRDIST 11-TEL 12-BAL
---	---

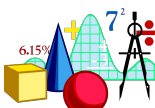
Select All Spread Zoom Select All Spread Zoom

Dependent variable list:

Independent variable list:

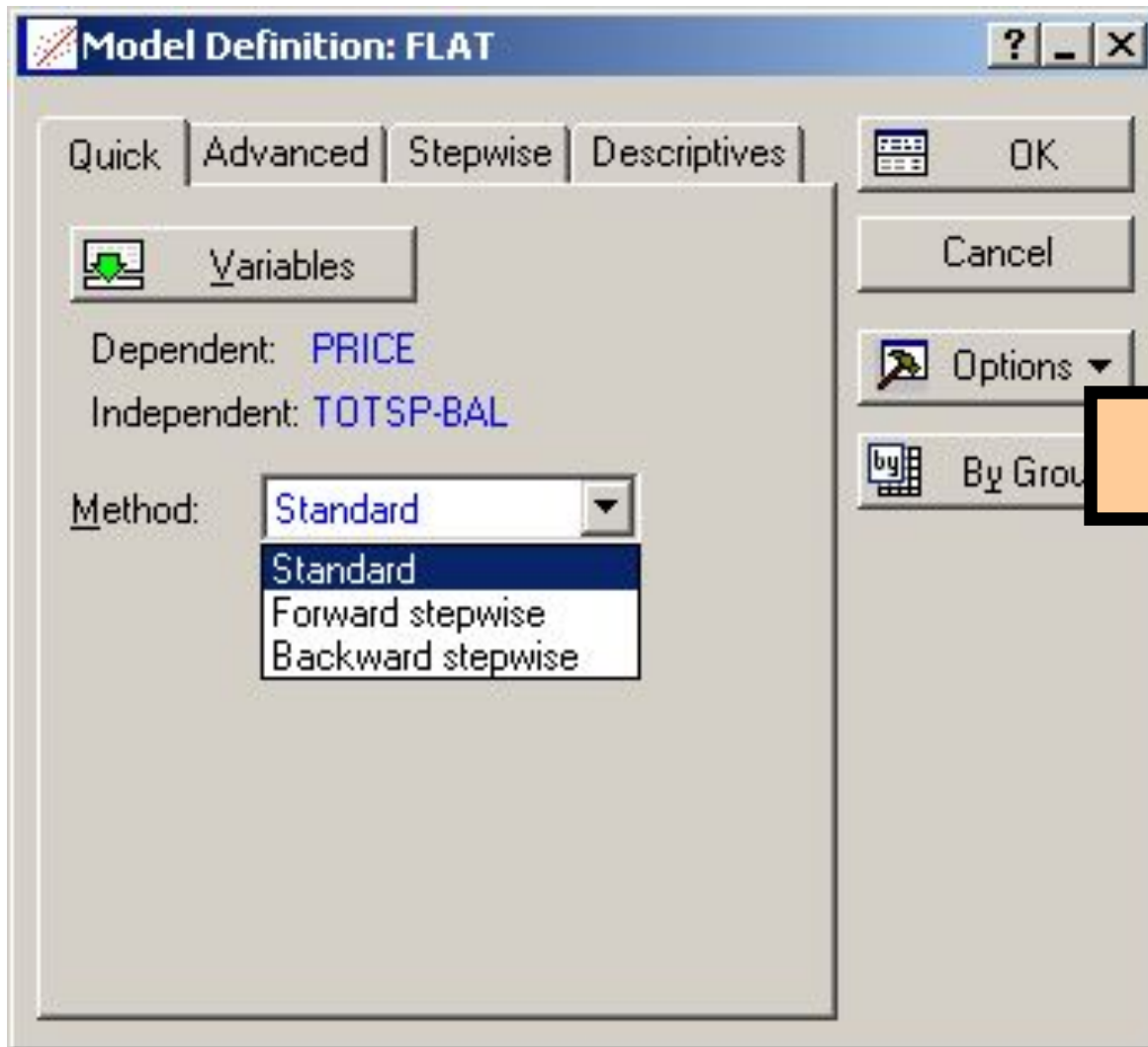
OK
Cancel

Выбор переменных

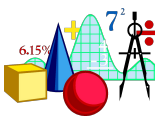




Начинаем анализ



Выбор метода





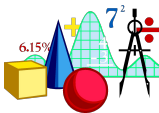
Выбор метода

В множественной линейной регрессии обычно реализовано три метода:

Standard – Стандартный

Forward stepwise – Прямой пошаговый метод

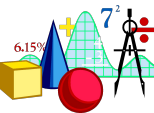
Backward stepwise - Обратный пошаговый метод





Выбор метода

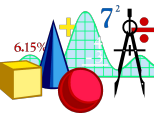
Standard – Стандартный –
включает в анализ сразу все
«независимые» переменные





Выбор метода

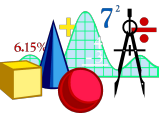
Forward stepwise – Прямой пошаговый метод – поочередно включает в регрессионное уравнение каждую переменную, начиная с наиболее тесно коррелирующей с зависимой переменной до тех пор, пока p -уровень значимости коэффициента b последней из включенных переменных не превысит заданное значение





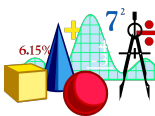
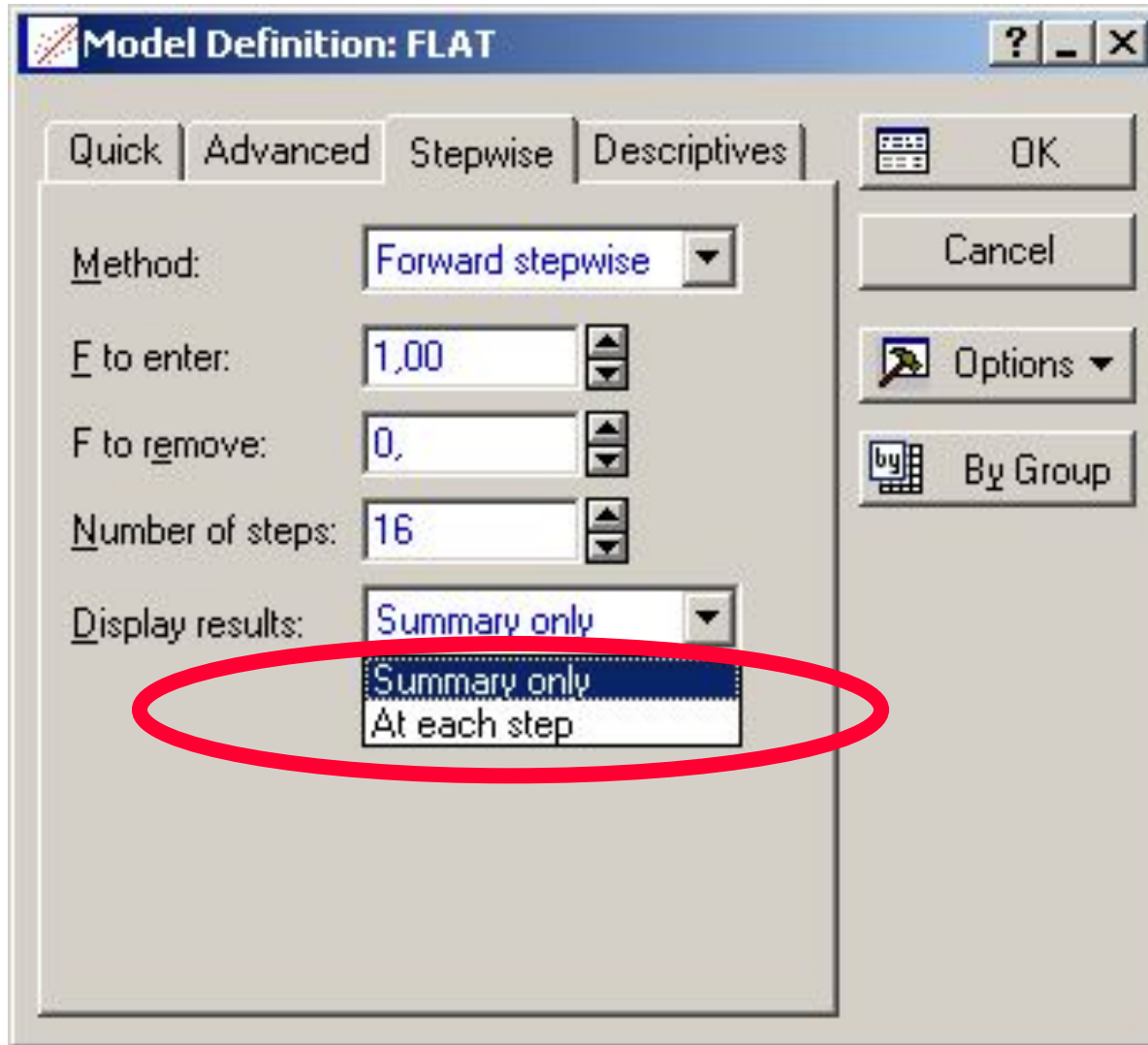
Выбор метода

Backward stepwise – обратный пошаговый метод – поочередно исключает переменные из анализа, начиная с той, которая имеет наибольшее значение p -уровня значимости коэффициента b , до тех пор, пока все оставшиеся переменные не будут иметь статистически значимые b -коэффициенты





Пошаговые методы



Начнем со стандартного метода

Multiple Regression Results: FLAT

Multiple Regression Results

Dependent: PRICE

No. of cases: 6286

Intercept: -2,568570470

Multiple R = ,73104735

R² = ,53443022

adjusted R² = ,53368828

Standard error of estimate: 5,19046111

Std. Error of Estimate (6275) = 5,19046111

F = 720,3109

df = 10,6275

p = 0,000000

PODSP beta=,323

DIST beta=-,31

FLOOR beta=,065

BAL beta=,007

LIVSP beta=,207

WALK beta=,089

METRDIST beta=-,12

KITSP beta=,322

BRICK beta=,1

TEL beta=,0

(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

Summary: Regression results

ANOVA (Overall goodness of fit)

Covariance of coefficients

Current sweep matrix

Partial correlations

Redundancy

Stepwise regression summary

ANOVA adjusted for mean

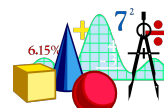
OK

Cancel

Options

By Group

Окно результатов





Итоги регрессии

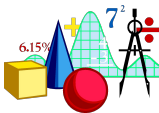
MULTIPLE R= ,73104735 RI= ,53443022 Adjusted RI= ,53368828
REGRESS. F(10,6275)=720,31 p<0,0000 Std.Error of estimate: 5,1905

N=6286	BETA	St. Err. of BETA	B	St. Err. of B	t(6275)	p-level
Intercpt			-2,56857	,723857	-3,5485	,000390
PODSP	,323351	,009384	,80850	,023462	34,4595	0,000000
LIVSP	,207418	,008804	,70594	,029965	23,5587	0,000000
KITSP	,321567	,009607	1,34235	,040104	33,4717	0,000000
DIST	-,306772	,009411	-,58142	,017836	-32,5980	0,000000
WALK	,088506	,008830	1,35457	,135147	10,0230	,000000
BRICK	,168090	,009589	2,91390	,166230	17,5293	0,000000
FLOOR	,064849	,008823	1,31784	,179308	7,3496	,000000
METRDIST	-,121001	,008892	-,1451	,015763	-13,6083	,000000
TEL	,047110	,008744	1,87622	,255430	5,3871	,000000
BAL	,006710	,008915	,13090	,173917	,7526	,451694

**Предсказательная
сила модели**

Коэффициенты

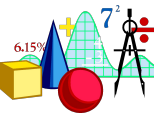
**Значимость
коэффициентов**





Анализ результатов

Переменная **Val (наличие балкона)
оказалась статистически незначима,
следовательно,
исключим ее из модели
и пересчитаем коэффициенты**





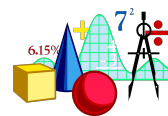
После исключения переменной **Val**

Regression Summary for Dependent Variable: PRICE (flat.sta)

MULTIPLE REGRESS. R= ,73101860 RI= ,53438819 Adjusted RI= ,53372049
F(9,6276)=800,34 p<0,0000 Std.Error of estimate: 5,1903

N=6286

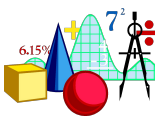
	BETA	St. Err. of BETA	B	St. Err. of B	t(6276)	p-level
Intercpt			-2,53787	,722681	-3,5117	,000448
PODSP	,323727	,009370	,80944	,023428	34,5495	0,000000
LIVSP	,207625	,008800	,70664	,029949	23,5944	0,000000
KITSP	,322584	,009511	1,34659	,039704	33,9160	0,000000
DIST	-,306952	,009407	-,58176	,017830	-32,6287	0,000000
WALK	,088259	,008824	1,35079	,135048	10,0023	,000000
BRICK	,168452	,009577	2,92018	,166015	17,5898	0,000000
FLOOR	,065837	,008725	1,33793	,177305	7,5459	,000000
METRDIST	-,120933	,008891	-,21439	,015762	-13,6019	,000000
TEL	,047307	,008739	1,38200	,255306	5,4131	,000000





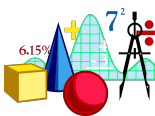
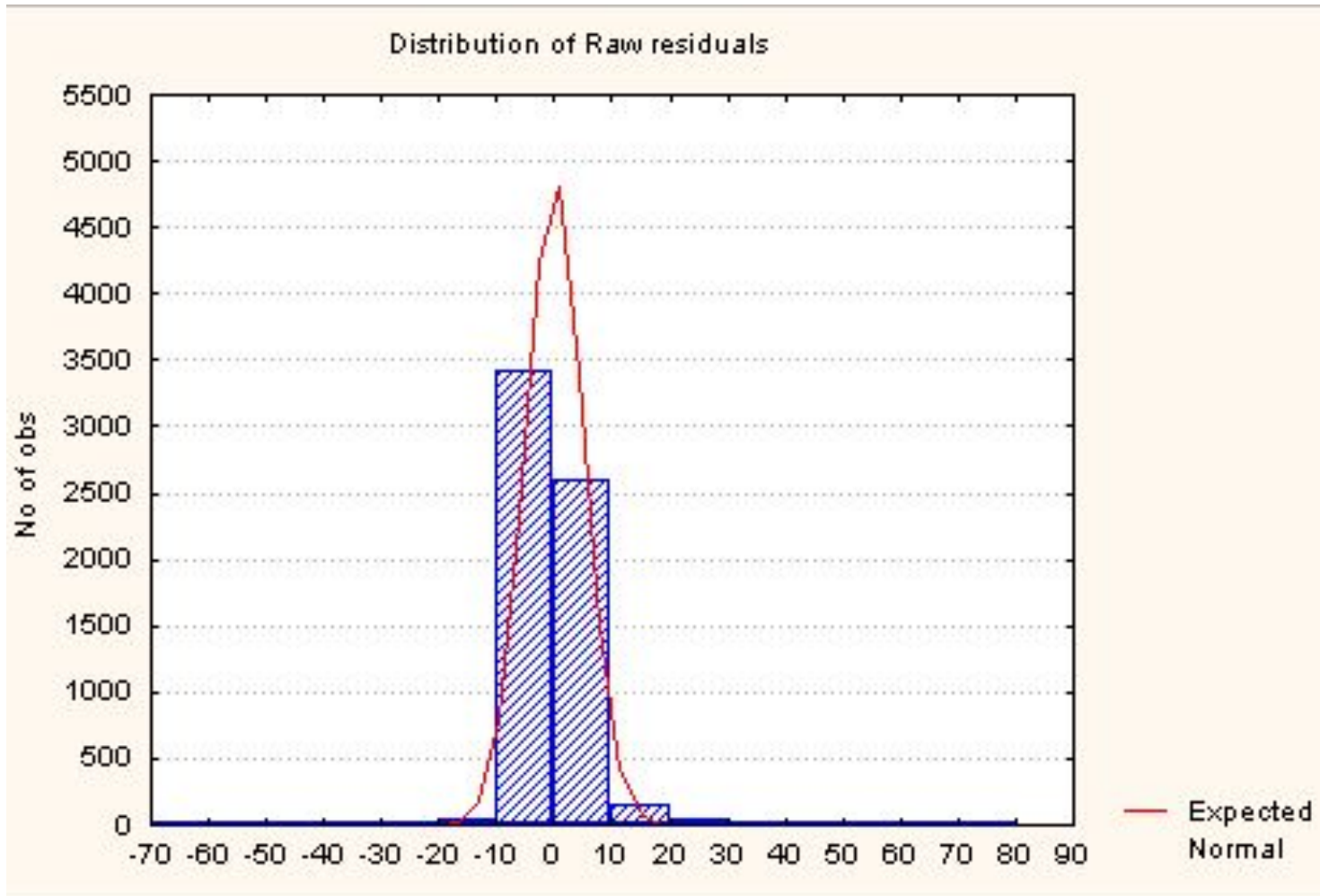
Теперь можно определить стоимость квартиры:

$$\begin{aligned} \text{Стоимость квартиры} = & 751 * \text{PODSP} + \\ & + 704 * \text{LIVSP} + 1290 * \text{KITSP} + \\ & + 20920 * \text{DIST_1} + 1300 * \text{WALK} + \\ & + 3256 * \text{BRICK} + 1282 * \text{FLOOR} + \dots \end{aligned}$$



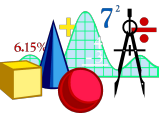
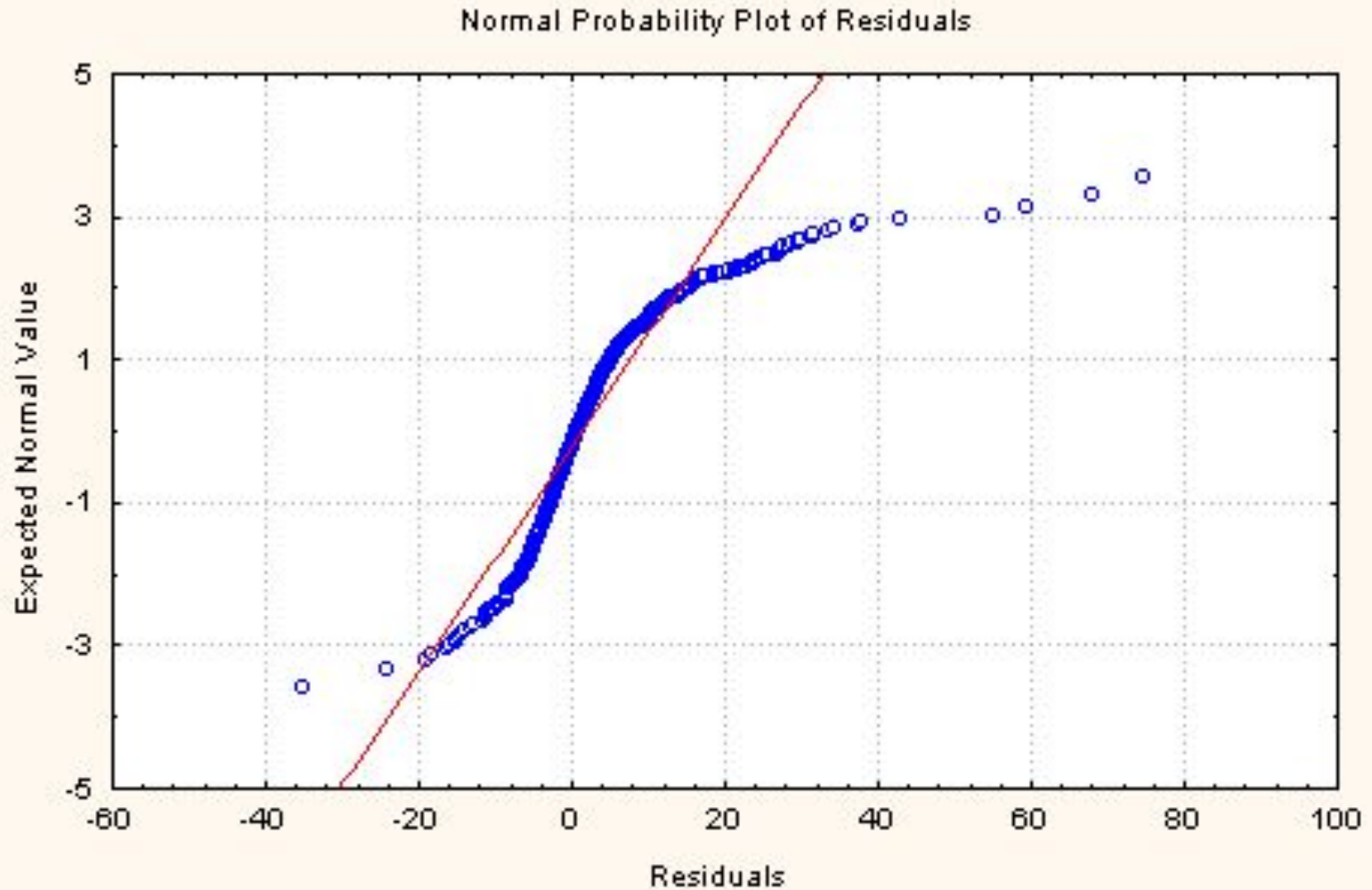


Оценим модель





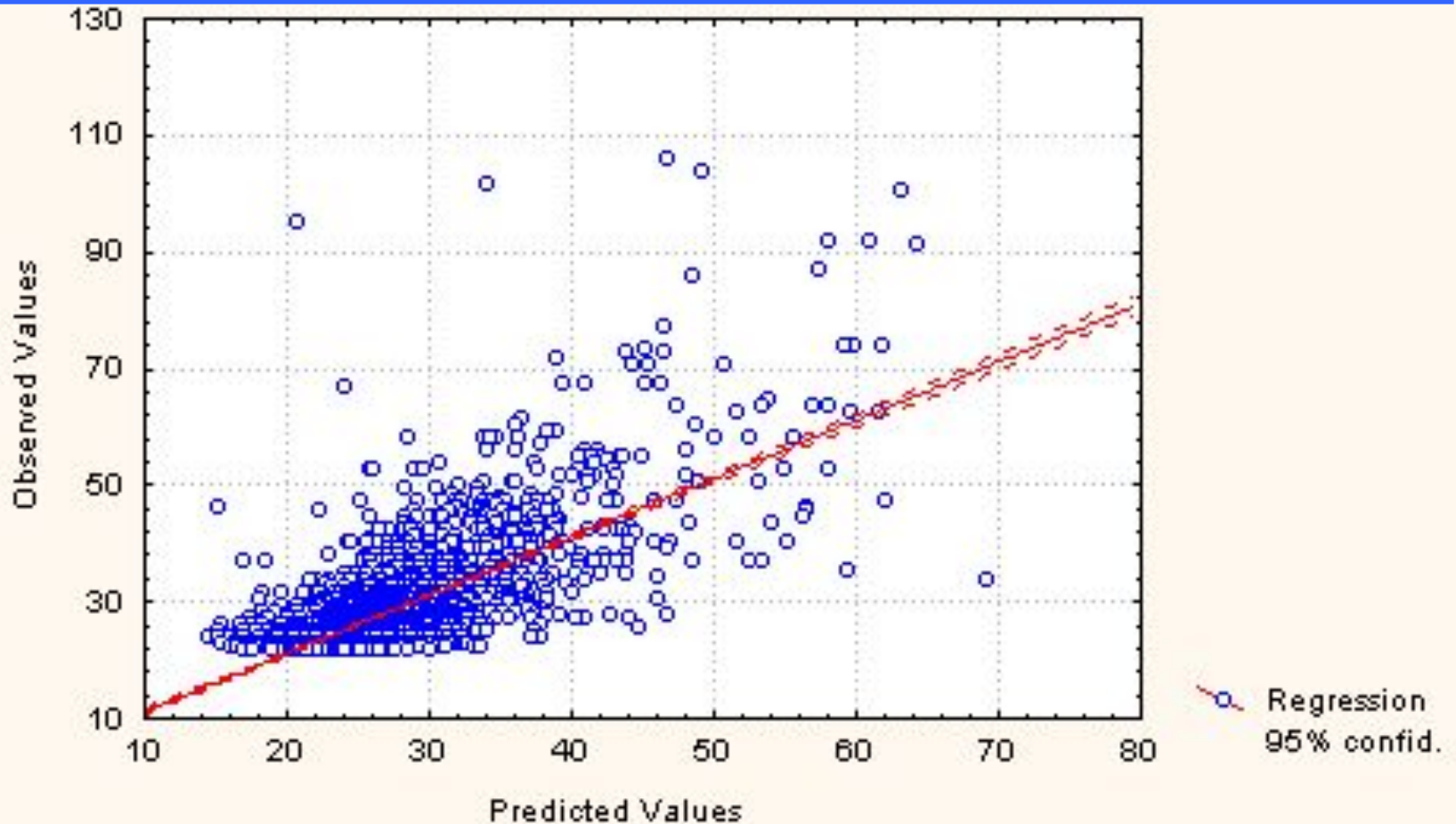
Оценим модель





Оценим модель

Коэффициент Дарбина-Ватсона=0,71

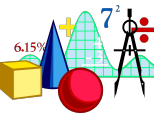




Интерпретация результатов

**На основе коэффициентов модели
можно сделать следующие выводы:**

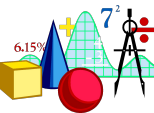
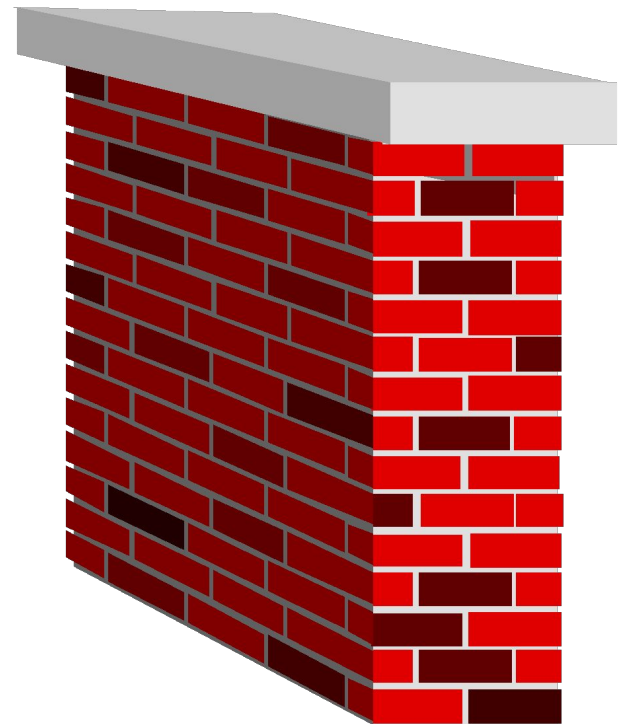
**Тот факт, что быстро добираться
до метро можно пешком, добавляет
к стоимости квартиры 1.300\$.**





Интерпретация результатов

**Тот факт, что тип
постройки
дома кирпичный,
а не панельный,
добавляет к стоимости
квартиры 3.200\$.
.... И Т.Д.**



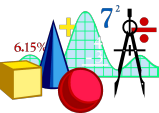


Интерпретация результатов

А.Д. Наследов (с.243):

«... знак β -коэффициента соответствует знаку коэффициента корреляции данной «независимой» и «зависимой» переменной.

Абсолютная величина β -коэффициента является максимальной – равна коэффициенту корреляции с зависимой переменной, если данная независимая переменная не коррелирует ни с одной из других независимых переменных»



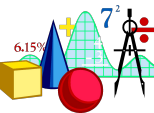


Пример 4 (реальные данные)

ЗП: ВР

НП:

- **согласованность (в %) отдельно для каждой группы**
- **Число альтернативных названий отдельно для каждой группы**
- **Субъективная зрительная сложность**
- **Частота употребления слова**

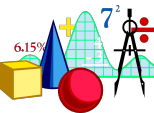




Пример 4

НП

- **Представляемость,**
- **Конкретность,**
- **Знакомость,**
- **Одушевленность**
- **Возраст, в котором слово выучено**
- **Длина слова (в фонемах)**

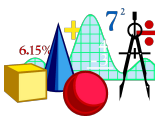




Пример 4

Корреляция между «знакомостью» и временем называния для трех групп:

Correlations (natika1.sta)	
BASIC STATS	Marked correlations are significant at $p < .05000$ N=228 (Casewise deletion of missing data)
Variable	FAMIL
A_PNRT	-,33
K1RT	-,36
K2RT	-,42





Пример 4

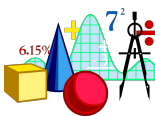
Результаты для группы 1:

Regression Summary for Dependent Variable: K1RT (natika1.sta)

MULTIPLE R= ,79373175 RI= ,63001010 Adjusted RI= ,61116802
REGRESS. F(11,216)=33,436 p<,00000 Std.Error of estimate: 190,43

N=228

	BETA	St. Err. of BETA	B	St. Err. of B	t(216)	p-level
Intercept			2426,180	439,5995	5,51907	,000000
SUBJVC	,132169	,052772	34,750	13,8747	2,50455	,013000
K1ALTN	,381729	,070274	34,962	6,4363	5,43201	,000000
K1CONS	-,333778	,076884	-4,170	,9606	-4,34131	,000022
PHONEME	,020833	,046628	3,629	8,1224	,44679	,655475
FREQ	,178617	,075980	54,880	23,3448	2,35083	,019632
IMAGEAB	-,147011	,060913	-127,393	52,7848	-2,41343	,016637
IMAGR	,047720	,047931	24,698	24,8070	,99560	,320557
CONCRETN	-,071831	,046534	-40,445	26,2013	-1,54361	,124147
FAMIL	-,034678	,082629	-34,714	82,7144	-,41968	,675135
ANIM	-,027722	,057646	-5,100	10,6043	-,48091	,631069
AOAVISUA	-,085443	,069790	-40,898	33,4051	-1,22429	,222176





Пример 2

Результаты для группы 2:

Regression Summary for Dependent Variable: K2RT (natika1.sta)

MULTIPLE REGRESS. R= ,78748446 RI= ,62013177 Adjusted RI= ,60078663
F(11,216)=32,056 p<,00000 Std.Error of estimate: 184,09

N=228

	BETA	St. Err. of BETA	B	St. Err. of B	t(216)	p-level
Intercpt			2994,943	423,4739	7,07232	,000000
SUBJVC	,071349	,053394	17,897	13,3927	1,33629	,182861
K2ALTN	,237817	,067380	38,878	11,0151	3,52948	,000509
K2CONS	-,312509	,071163	-5,074	1,1555	-4,39147	,000018
PHONEME	,046083	,047243	7,658	7,8511	,97546	,330426
FREQ	-,136042	,077675	-39,877	22,7679	-1,75144	,081290
IMAGEAB	-,093817	,061664	-77,559	50,9783	-1,52142	,129617
IMAGR	-,172530	,048187	-85,187	23,7923	-3,58046	,000424
CONCRETN	-,032461	,046268	17,487	24,8533	-,70159	,483687
FAMIL	,019014	,083604	18,159	79,8416	,22744	,820301
ANIM	-,070173	,058620	-12,316	10,2877	-1,19712	,232574
AOAVISUA	-,152400	,066030	-69,592	30,1521	-2,30803	,021944





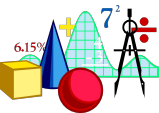
**И что же
делать?!!**





Будь бдительным!

**Так смело можно интерпретировать
регрессионные к-ты только если
независимые переменные действительно
независимы – не коррелируют друг с
другом!**

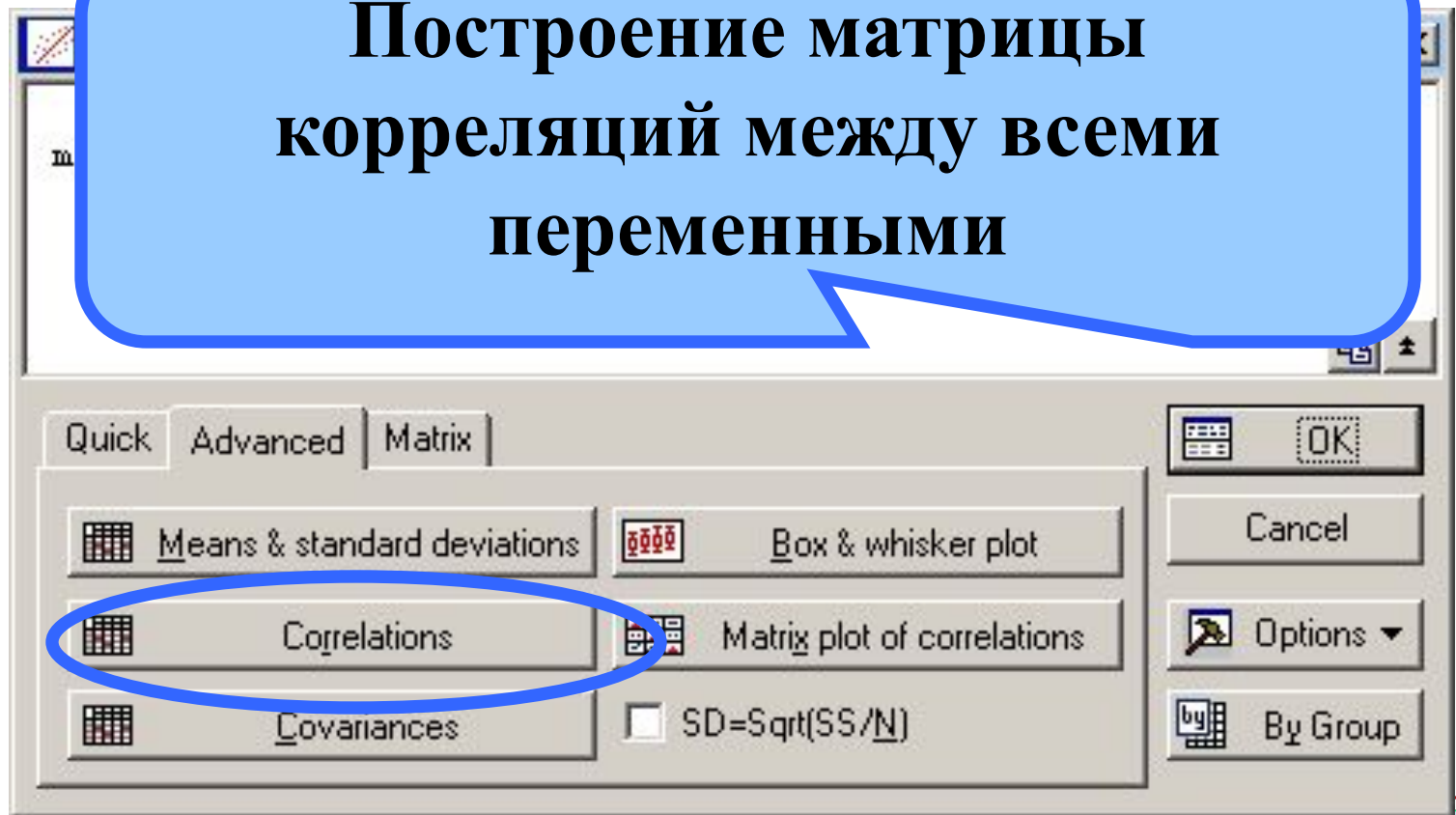




Будь бдительным!

Для проверки возможных связей между НП в программе STATISTICA есть много возможностей

Построение матрицы корреляций между всеми переменными





Будь бдительным!

Для примера 4 матрица корреляций имеет вид:

STATISTICA: Multiple Regression - [Correlations (natika1.sta)]

File Edit View Analysis Graphs Options Window Help

-.410479049906374 Columns Rows

MULTIPLE REGRESS.	SUBJVC	K2ALTN	K2CONS	PHONEME	FREQ	IMAGEAB	IMAGR	CONCRET
SUBJVC	1,000000	,269063	-,245574	,148097	-,319400	-,123795	,042718	-,021884
K2ALTN	,269063	1,000000	-,768943	,125825	-,330492	-,245171	-,127735	-,112088
K2CONS	-,245574	-,768943	1,000000	-,204411	,425586	,323533	,086221	,101269
PHONEME	,148097	,125825	-,204411	1,000000	-,108482	-,004907	-,062641	,109740
FREQ	-,319400	-,330492	,425586	-,108482	1,000000	,583707	-,238037	,188913
IMAGEAB	-,123795	-,245171	,323533	-,004907	,583707	1,000000	,115890	,375789
IMAGR	,042718	-,127735	,086221	-,062641	-,238037	,115890	1,000000	,151932
CONCRETN	-,021884	-,112088	,101269	,109740	,188913	,375789	,151932	1,000000
FAMIL	-,410479	-,324526	,401543	-,103141	,809196	,598951	-,173359	,185974
ANIM	,513310	,147617	-,071660	-,107710	-,301370	-,025320	,242306	-,068021
AOAVISUA	-,172120	-,385994	,501935	-,372637	,516567	,467924	,017702	,082706
K2RT	,249717	,639049	-,692563	,235305	-,437442	-,432037	-,236111	-,178731

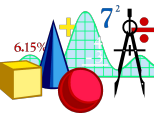
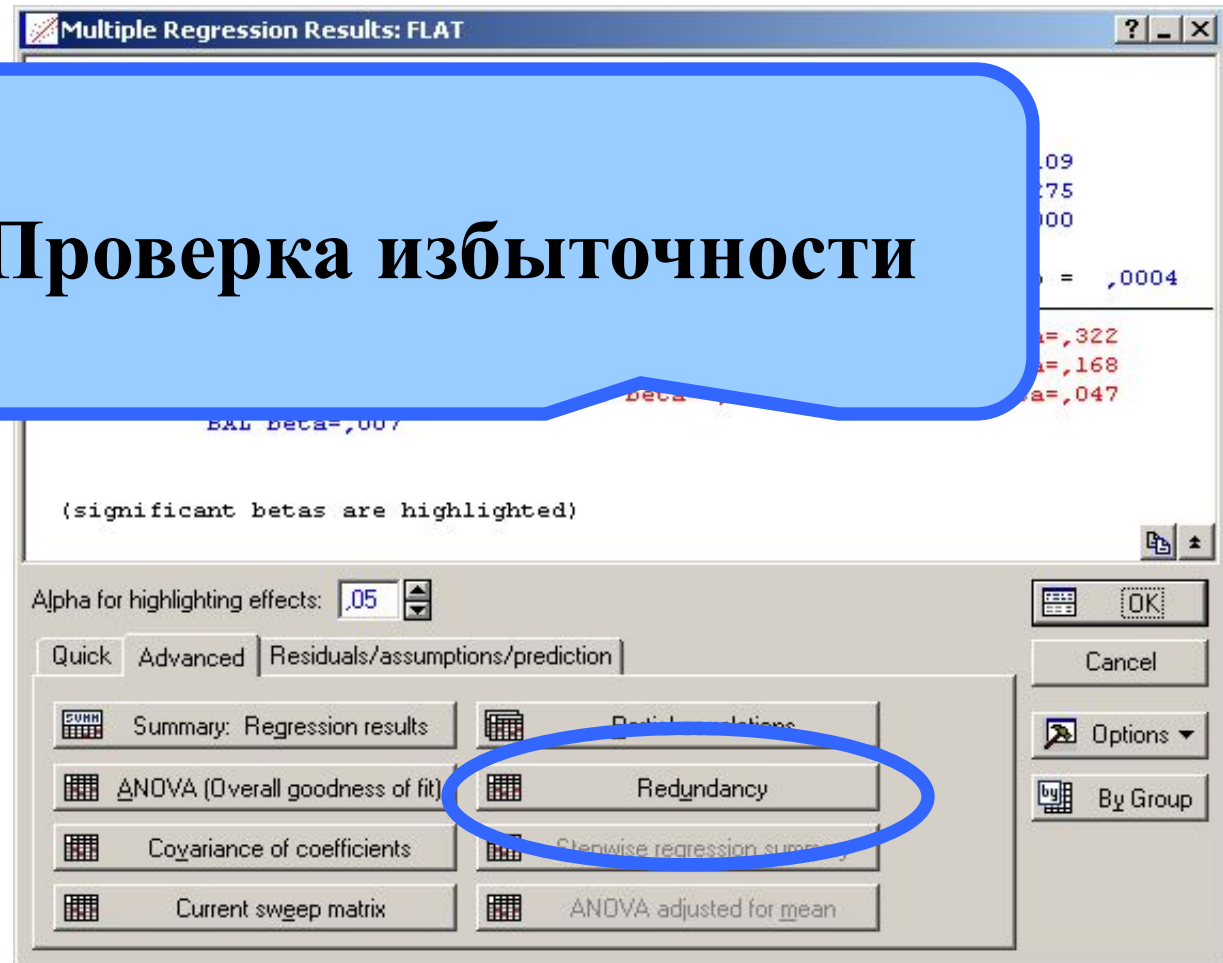




Будь бдительным!

Для проверки возможных связей между НП в программе STATISTICA есть много возможностей

Проверка избыточности



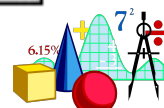


Будь бдительным!

Для проверки возможных связей между НП в программе STATISTICA

Чем меньше толерантность переменной, тем больше ее избыточность (т.е. тем больше она коррелирует с другими переменными)

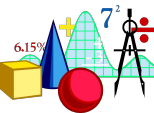
MULTIPLE REGRESS.	R-square column contains variable with all other variables			
variable	Toleran.	R-sq		
SUBJVC	.616877	.38		
K2ALTN	.387360	.61		
K2CONS	.347275	.65		
PHONEME	.787960	.2120		.06622
FREQ	.291488	.708512	-.118333	-.073449
IMAGEAB	.462499	.537501	-.102969	-.063803
IMAGR	.757404	.242596	-.236697	-.150151
CONCRETN	.021519	.178481	-.047683	-.029422
FAMIL	.251611	.748389	.015473	.009538
ANIM	.511782	.488218	-.081184	-.050203
AOAVISUA	.403360	.596640	-.155140	-.096790





Будь бдительным!

**Проверяйте наличие корреляций
между независимыми переменными
и используйте пошаговые методы
множественной линейной регрессии**



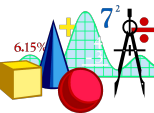


Шкалы наименований

**В примере 3 использовались
дихотомические шкалы.**

**А что делать, если попалась шкала
наименований?**

**Не спешите расстраиваться! Надо ее
просто перекодировать!**





Шкалы наименований

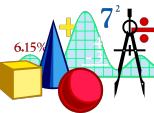
Если есть шкала «профессия» с кодами

1 – клерк

2 – охранник

3 – менеджер

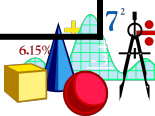
то перекодируем ее в 3 переменных!






Шкалы наименований

профессия		клерк	охранник	менеджер
1		1	0	0
Теперь смело можно проводить множественный регрессионный анализ!				
...
3		0	0	1





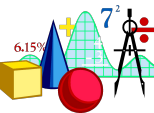
**Мне кажется, Вы
уже достаточно
регрессировали...**






**К практическому занятию по регрессионному анализу
надо прочитать:**

- **Нестеренко А.И. и др. Прогноз тревожности у студенток на основании их типологических различий// ПЖ, 2003, т.24, № 6, с. 37-46**
- **Нечаева Е.С., Козубовский В.М. Ошибки интерпретации регрессионных моделей в психологических исследованиях// ПЖ (белорусский), 2006, т.26, № 2, с. 82-85**

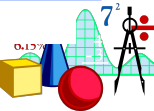




**А что делать, если
зависимая
переменная не
количественная, а
качественная?**



**Можно променять
ДИСКРИМИНАНТНЫЙ
АНАЛИЗ!**





СПАСИБО
ЗА
ВНИМАНИЕ!

