

Использование прикладного ПО для построения линейной регрессионной модели

1. Использование MS Excel для построения линейной регрессионной модели

На отдельном листе вводятся и оформляются исходные данные

Microsoft Excel - Пример 1 (откр лекция)

Введите вопрос

100%

Аrial Cyr 10 Ж К Ч

H23 fx

По территориям региона приводятся данные за 199X г.

	А	В	С	Д	Е	Ф	Г
1	По территориям региона приводятся данные за 199X г.						
2							
3	Номер региона	Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., х	Среднедневная зарплата, руб., у				
4	1	78	133				
5	2	82	148				
6	3	87	134				
7	4	79	154				
8	5	89	162				
9	6	106	195				
10	7	67	139				
11	8	88	158				
12	9	73	152				
13	10	87	162				
14	11	76	159				
15	12	115	173				
16							

Исх данные / Лист2 / Лист3 /

Готово

NUM

Рекомендация:
в дальнейшем на эти данные только ссылаться, а не дублировать их ввод с клавиатуры

Предварительный анализ

1. Построить корреляцию визуальный анализ

- Выделить диапазон
(2 столбца данных)
- Вызвать мастер диаграмм
- Выбрать точечную диаграмму

	А	В	С	Д	Е	Ф	Г	Н	І	Ј
1	По территориям региона приводятся данные за 199Х г.									
2										
3	Номер региона	Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., х	Среднедневная зарплата, руб., у							
4	1	78	133							
5	2	82	148							
6	3	87	134							
7	4	79	154							
8	5	89	162							
9	6	106	195							
10	7	67	139							
11	8	88	158							
12	9	73	152							
13	10	87	162							
14	11	76	159							
15	12	115	173							
16										
17										
18										
19										
20										
21										


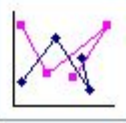



Мастер диаграмм (шаг 1 из 4): тип диаграммы

Стандартные Нестандартные

Тип:

- Гистограмма
- Линейчатая
- График
- Круговая
- Точечная
- С областями
- Кольцевая
- Лепестковая
- Поверхность
- Пузырьковая

Вид:



Точечная диаграмма позволяет сравнить пары значений.

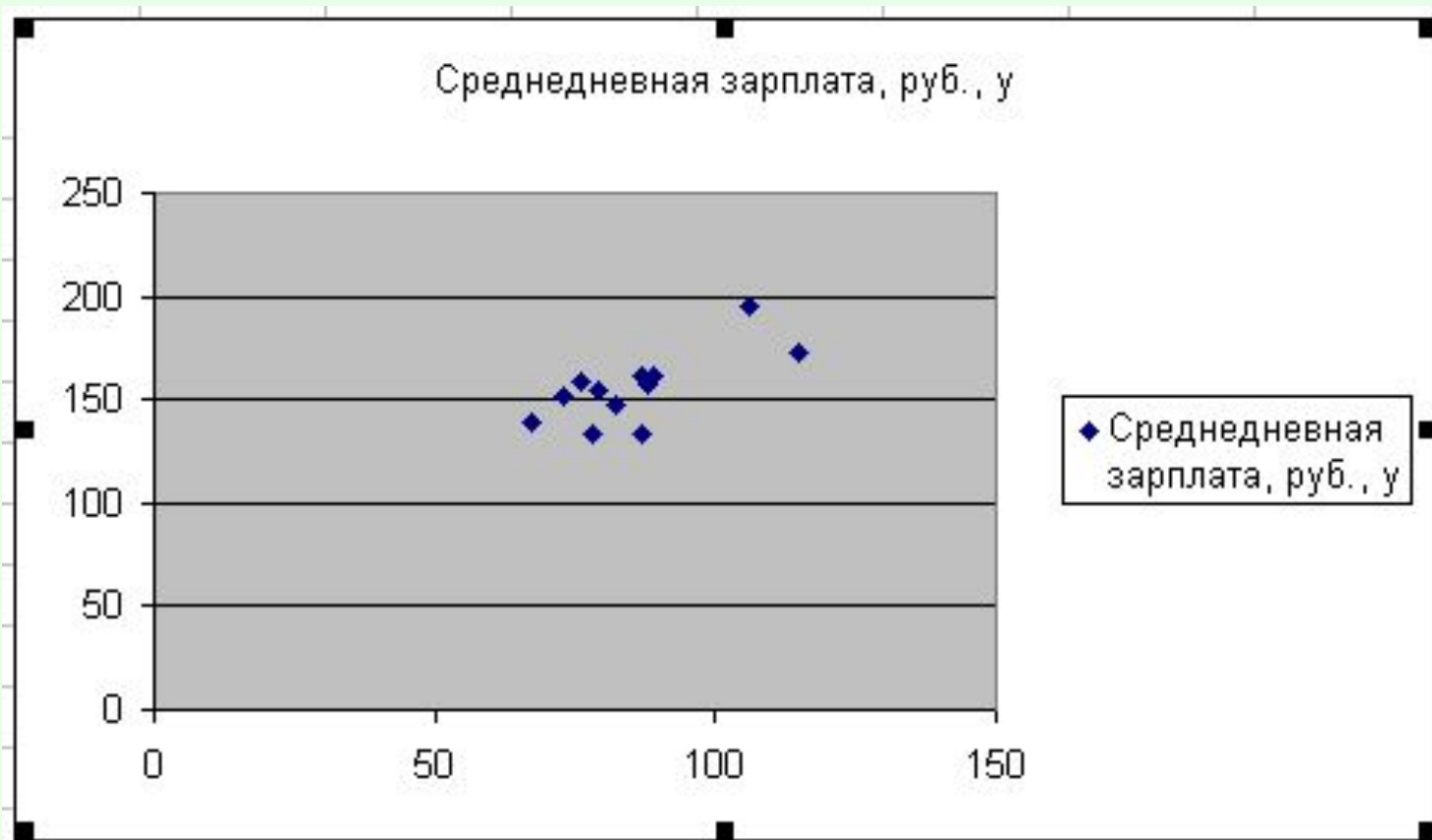
Просмотр результата

Отмена < Назад Далее > Готово

Предварительный анализ

1. Построить корреляционное поле – визуальный анализ

Нажать кнопку «Готово» и
появится диаграмма –
корреляционное поле



Предварительный анализ

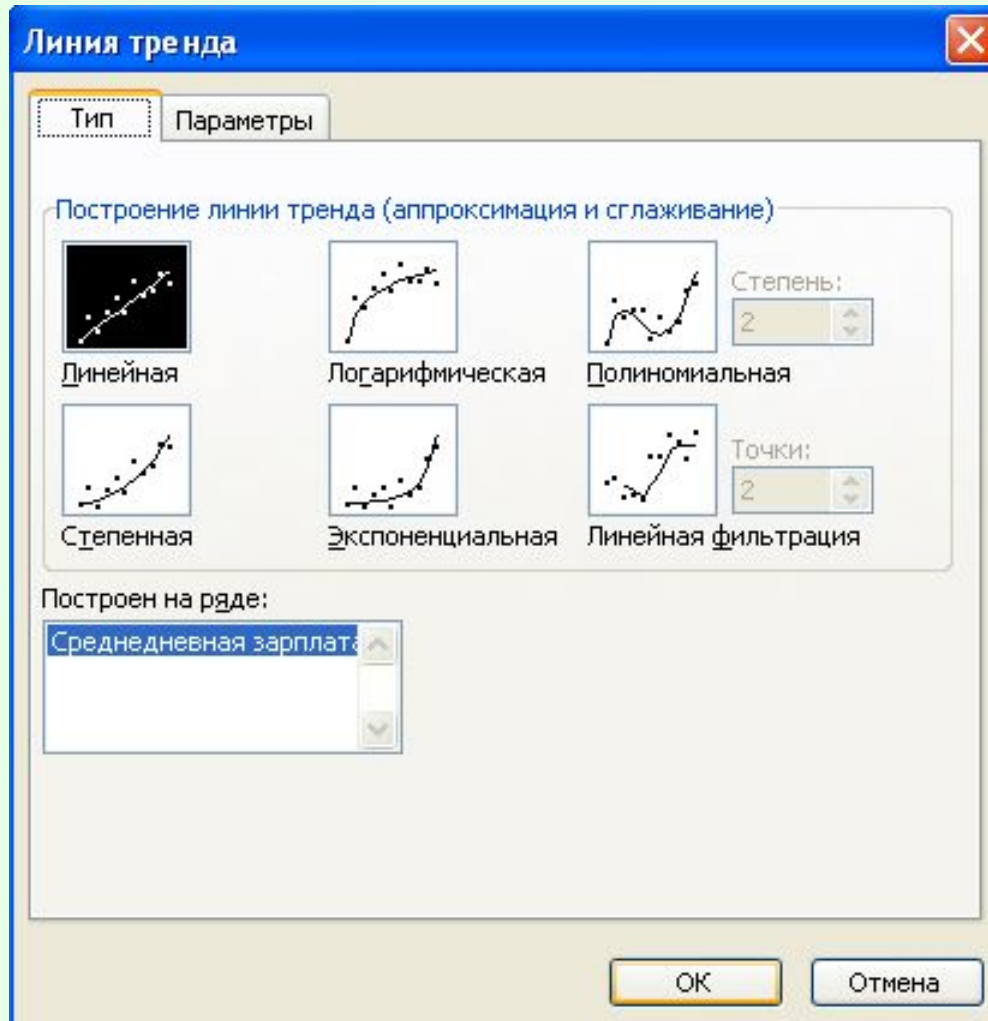
2. Добавить линию тренда, показать его уравнение и коэффициент детерминации

1. Правый щелчок по точкам на диаграмме
2. В контекстном меню выбрать: **Добавить линию тренда ...**



Предварительный анализ

2. Добавить линию тренда, показать его уравнение и коэффициент детерминации



1. Появится окно «ЛИНИЯ ТRENDA»
2. На вкладке «ТИП» выбрать «ЛИНЕЙНАЯ» (при необходимости и другие)

Предварительный анализ

2. Добавить линию тренда, показать его уравнение и коэффициент детерминации

Линия тренда

Тип Параметры

Название аппроксимирующей (сглаженной) кривой

☒ автоматическое: Линейный (Среднедневная зарплата, руб., y)

☐ другое:

Прогноз

вперед на: 0 единиц

назад на: 0 единиц

☐ пересечение кривой с осью Y в точке: 0

☒ показывать уравнение на диаграмме

☒ поместить на диаграмму величину достоверности аппроксимации (R^2)

OK Отмена

На вкладке «ПАРАМЕТРЫ»
установить флажки:

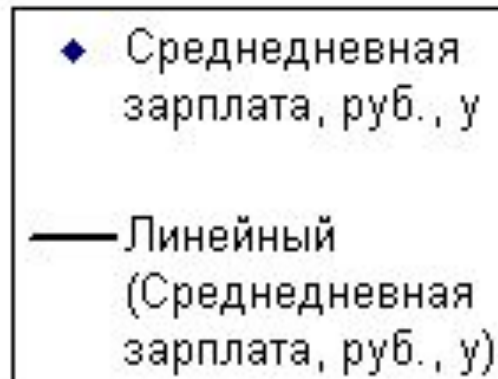
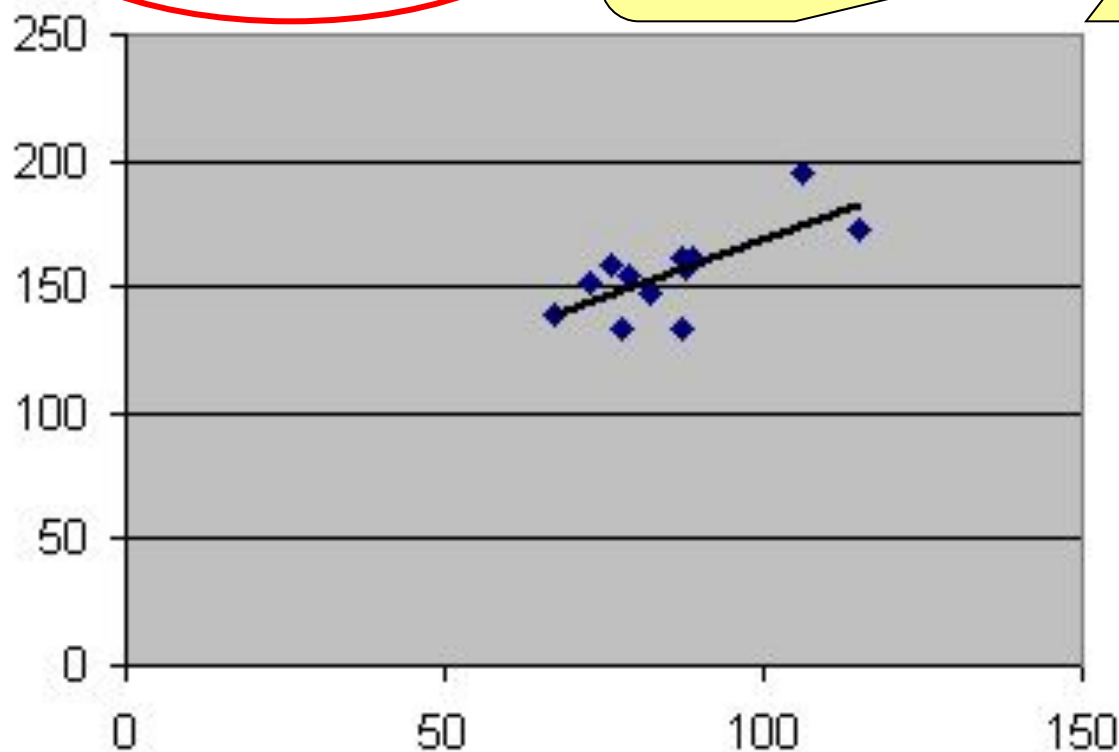
- Показывать уравнение на диаграмме
- Поместить на диаграмму величину достоверности аппроксимации (коэффициент детерминации)

Предварительный анализ

2. Добавить линию тренда, показать его уравнение и коэффициент детерминации

$$y = 0,9204x + 76,976$$
$$R^2 = 0,5199$$

Уравнение линейной регрессии
Коэффициент детерминации



Предварительный анализ

Демонстрация предварительного анализа в MS Excel

1. Выделение и использование области исходных данных
2. Построение диаграммы
3. Добавление линии тренда, уравнения и коэффициента детерминации



Парная линейная регрессия в MS Excel

1) Использование диаграммы

1. Ввести данные в 2 столбца
2. Построить точечную диаграмму по данным
3. Добавить на диаграмму линию тренда и показать его уравнение

Линейный регрессионный анализ

Используется встроенная статистическая функция **ЛИНЕЙН**, которая определяет параметры линейной регрессии

Мастер функций - шаг 1 из 2

Поиск функции:

Введите краткое описание действия, которое нужно выполнить, и нажмите кнопку "Найти"

Найти

Категория: Статистические

Выберите функцию:

- КОРРЕЛ
- КРИТБИНОМ
- ЛГРФПРИБЛ
- ЛИНЕЙН**
- ЛОГНОРМОБР
- ЛОГНОРМРАСП
- МАКС

ЛИНЕЙН(известные_значения_y;известные_значения_x;...)
Возвращает параметры линейного приближения по методу наименьших квадратов.

[Справка по этой функции](#)

ОК Отмена

1. Меню ВСТАВКА – пункт ФУНКЦИЯ
2. В окне МАСТЕР ФУНКЦИЙ:
 - Категория СТАТИСТИЧЕСКИЕ
 - В списке выбрать ЛИНЕЙН


Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН

Аргументы функции


ЛИНЕЙН

Известные_значения_y




= ссылка

Известные_значения_x




= ссылка

Конст



= логическое

Статистика



= логическое

=

Возвращает параметры линейного приближения по методу наименьших квадратов.

Известные_значения_y:

множество фактических значений переменной y

=

[Справка по этой функции](#)

Значение:

OK

Отмена


Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН

Аргументы функции


ЛИНЕЙН

Известные_значения_y




= ссылка

Известные_значения_x




= ссылка

Конст



= логическое

Статистика



= логическое

=

Возвращает параметры линейного приближения по методу наименьших квадратов.

Известные_значения_x:
множество фактических значений переменной x

=

[Справка по этой функции](#)

Значение:

OK





Отмена

Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН

Аргументы функции

ЛИНЕЙН

Известные_значения_y	<input type="text"/>		= ссылка
Известные_значения_x	<input type="text"/>		= ссылка
Конст	<input type="text"/>		= логическое
Статистика	<input type="text"/>		= логическое

Конст: логическое значение:

Истина (1): параметр а (свободный член) вычисляется обычным образом

Ложь (0): параметр а (свободный член) равен нулю

Уравнение линейной парной регрессии:





$$y = a + bx$$

Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН

Аргументы функции

ЛИНЕЙН

Известные_значения_y	<input type="text"/>		= ссылка
Известные_значения_x	<input type="text"/>		= ссылка
Конст	<input type="text"/>		= логическое
Статистика	<input type="text"/>		= логическое

Статистика: логическое значение:
Истина (1): требуется показать дополнительную статистику по регрессии
Ложь (0): не требуется показать дополнительную статистику

[Справка по этой функции](#) Значение:

Линейный регрессионный анализ

Результаты функции **ЛИНЕЙН**

	A	B	C	D	E	F	G	H	I	J	K	L
1	Результаты линейного регрессионного анализа											
2												
3		Значение коэффициента b						Значение коэффициента a				
4		Среднеквадратическое отклонение b						Среднеквадратическое отклонение a				
5		Коэффициент детерминации R^2						Среднеквадратическое отклонение y				
6		F-статистика						Число степеней свободы				
7		Регрессионная сумма квадратов (RSS)						Остаточная сумма квадратов (ESS)				
8												

Результат выводится в диапазоне 5*2 (5 строк; 2 столбца)

Уравнение линейной парной регрессии:

$$y = a + bx$$

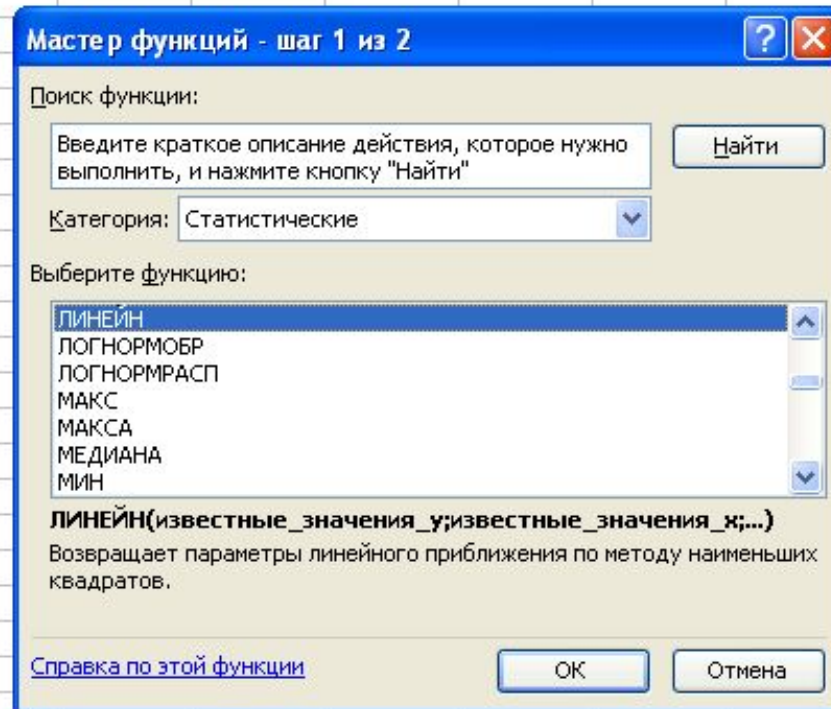
Линейный регрессионный анализ

Ход ввода параметров функции **ЛИНЕЙН**

Результаты линейного регрессионного анализа

Значение коэффициента b	=		Значение коэффициента a
Среднеквадратическое отклонение b			Среднеквадратическое отклонение a
Коэффициент детерминации R^2			Среднеквадратическое отклонение y
F-статистика			Число степеней свободы
Регрессионная сумма квадратов (RSS)			Остаточная сумма квадратов (ESS)

1. Выделить диапазон ячеек 5*2
2. Вызвать окно вставки функции (ВСТАВКА-ФУНКЦИЯ-СТАТИСТИЧЕСКИЕ-ЛИНЕЙН)







Линейный регрессионный анализ

Ход ввода параметров функции ЛИНЕЙН

Аргументы функции

ЛИНЕЙН

Известные_значения_y	<input type="text"/>		= ссылка
Известные_значения_x	<input type="text"/>		= ссылка
Конст	<input type="text"/>		= логическое
Статистика	<input type="text"/>		

Возвращает параметры линейного приближения по методу наименьших квадратов

3. Появится окно ввода параметров функции

значений y, для которых уже
 $mx + b$.

Нажать эту кнопку для выделения диапазона значений y

[Справка по этой функции](#) Значение:

Линейный регрессионный анализ

Ход ввода параметров функции **ЛИНЕЙН**

4. На листе «Исх данные» выделить диапазон C4:C15







Линейный регрессионный анализ

Ход ввода параметров функции ЛИНЕЙН

Аргументы функции

ЛИНЕЙН

Известные_значения_y	'Исх данные'!C4:C15		= {133;148;134;154;1
Известные_значения_x	'Исх данные'!B4:B15		= {78;82;87;79;89;10
Конст	1		= ИСТИНА
Статистика	1		= ИСТИНА

= {0,920430552680604;76

Возвращает параметры линейного приближения по методу наименьших квадратов.

5. Аналогично ввести диапазон значений x

6. В остальные поля ввести 1 (истина)

7. Нажать Ок

[Справка по этой функции](#) Значение: 0,920430553

Линейный регрессионный анализ

Ход ввода параметров функции **ЛИНЕЙН**

	A	B	C	D	E	F	G	H
1	Результаты линейного регрессионно							
2								
3		Значение коэффициента b				0,9204		Зна
4		Среднеквадратическое отклонение b						Сре
5		Коэффициент детерминации R^2						Сре
6		F-статистика						Чис
7		Регрессионная сумма квадратов (RSS)						Ост
8								

Появится **только** значение параметра b

Линейный регрессионный анализ

Результаты линейного регрессионного анализа					
Значение коэффициента b		0,9204	76,976	Значение коэффициента a	
Среднеквадратическое отклонение b		0,2797	24,212	Среднеквадратическое отклонение a	
Коэффициент детерминации R^2		0,5199	12,55	Среднеквадратическое отклонение y	
F-статистика		10,828	10	Число степеней свободы	
Регрессионная сумма квадратов (RSS)		1705,3	1574,9	Остаточная сумма квадратов (ESS)	

0,9204	76,976
0,2797	24,212
0,5199	12,55
10,828	10
1705,3	1574,9

Результаты
предварительного
анализа

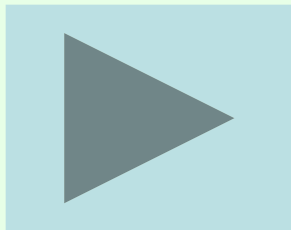
$$y = 0,9204x + 76,976$$

$$R^2 = 0,5199$$

Предварительный анализ

Демонстрация анализа с помощью функции ЛИНЕЙН в MS Excel

1. Вставка функции, ввод параметров, вывод результатов
2. Дальнейшее использование результатов, например, для расчета RSS
3. Автоматический пересчет значений функции после изменения исходных данных



Парная линейная регрессия в MS Excel

2) Использование функции ЛИНЕЙН

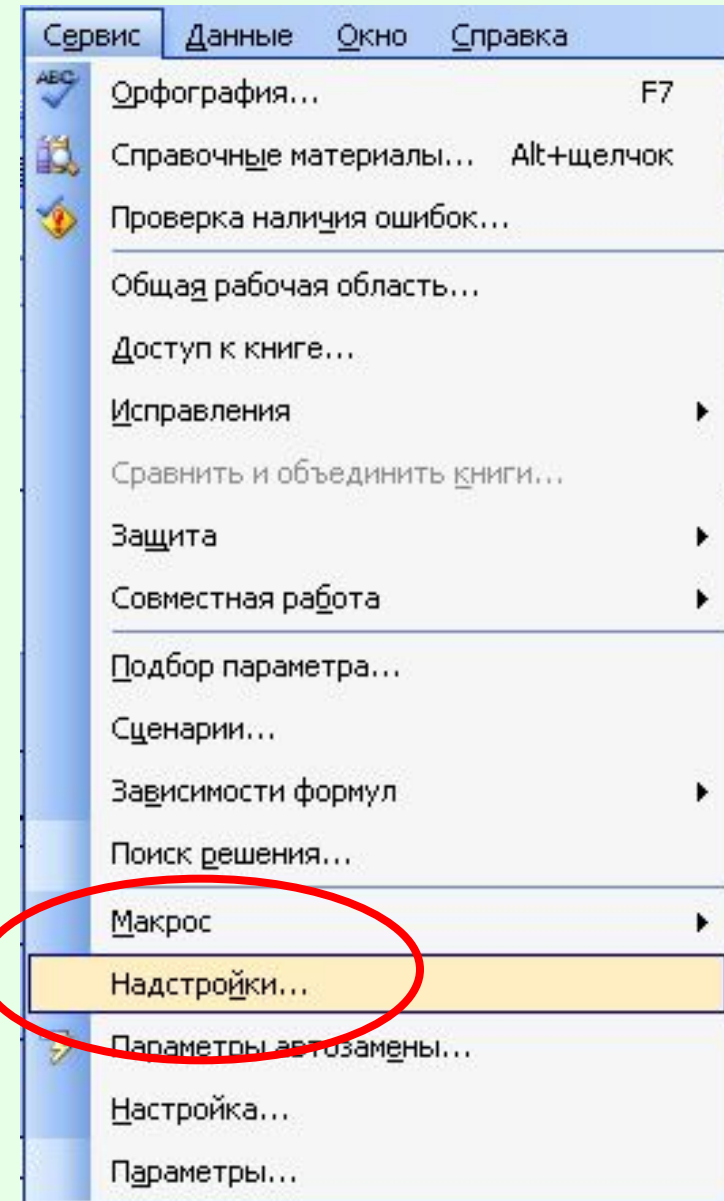
1. Ввести данные в 2 столбца
2. Выделить диапазон свободных ячеек 5x2
3. Вызвать функцию ЛИНЕЙН и ввести ее параметры
4. Нажать F2, а затем CTRL+SHIFT+ENTER

Линейный регрессионный анализ

Использование инструмента анализа данных РЕГРЕССИЯ

1. Проверить доступ к пакету анализа

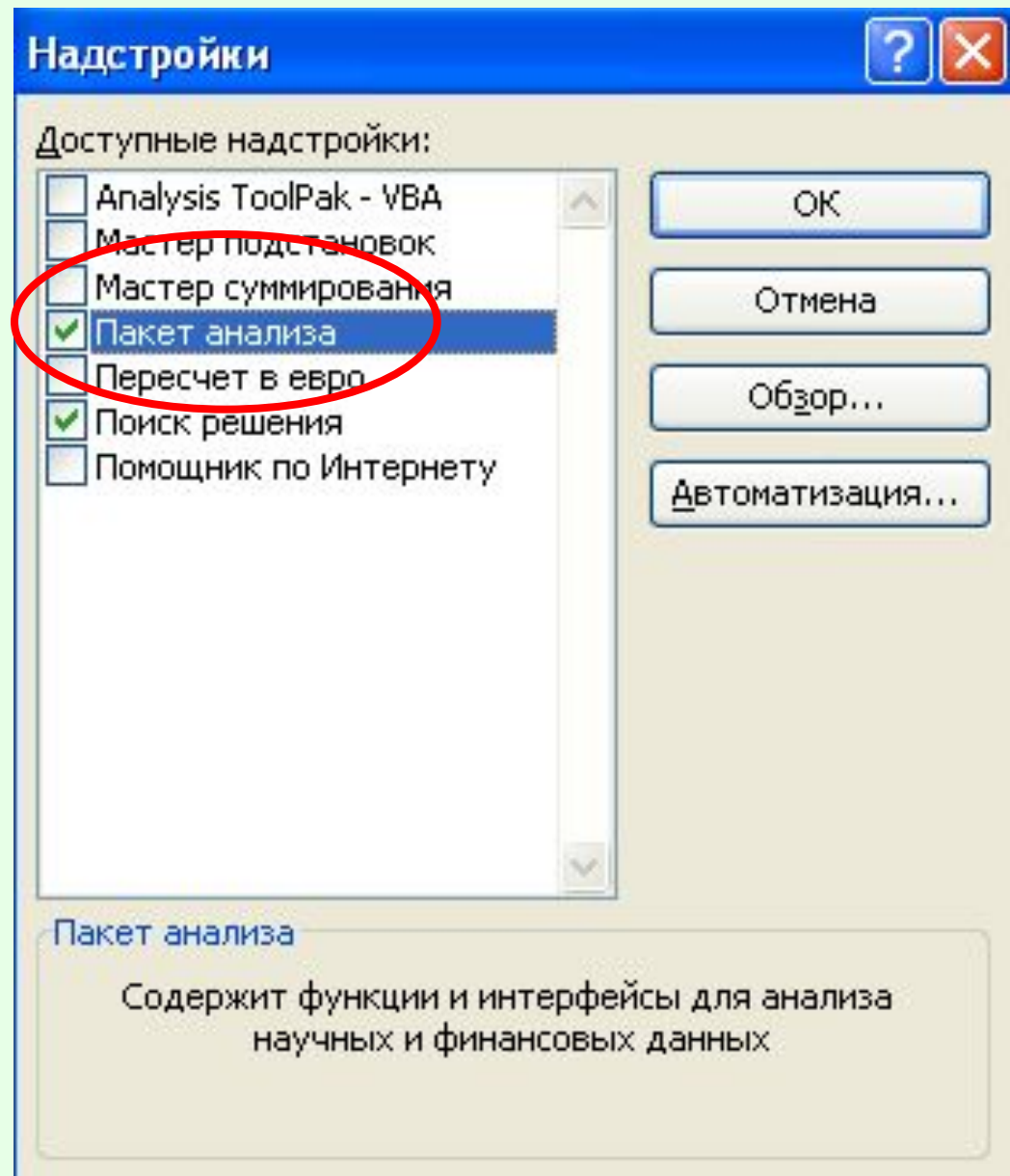
В главном меню последовательно выбрать
СЕРВИС / НАДСТРОЙКИ.



Линейный регрессионный анализ

2. Установить флажок
ПАКЕТ АНАЛИЗА

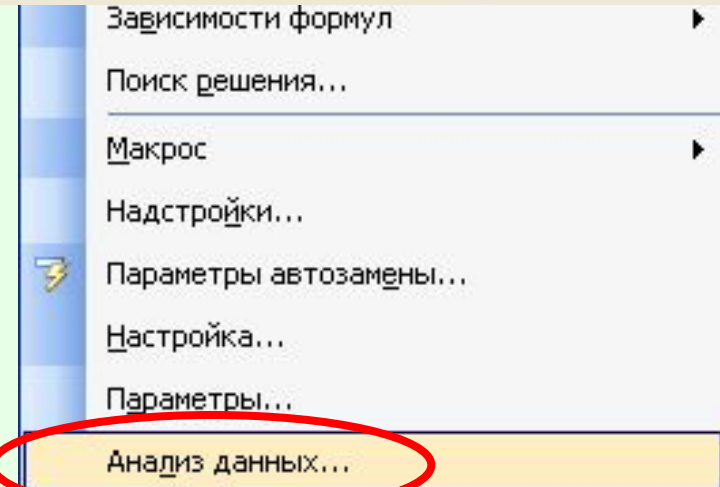
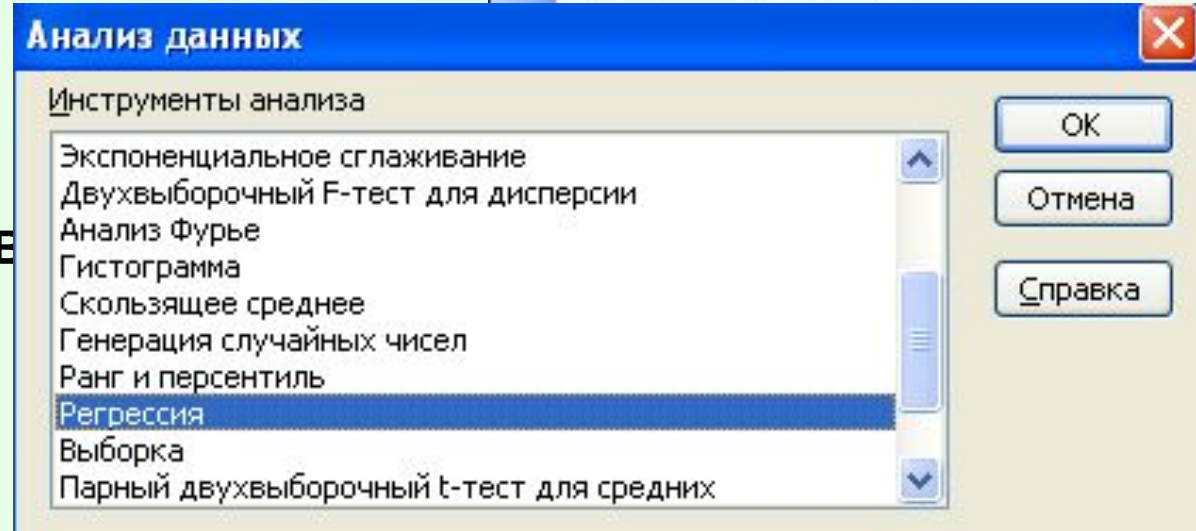
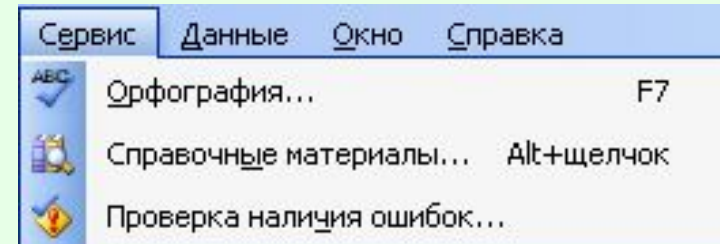
3. Нажать **Ок**



Линейный регрессионный анализ

Использование инструмента
анализа данных РЕГРЕССИЯ

4. В главном меню
СЕРВИС/
АНАЛИЗ ДАННЫХ/
РЕГРЕССИЯ



Использование инструмента РЕГРЕССИЯ

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

Константа - ноль ☐

Остатки

☐ Остатки ☐ График остатков

☐ Стандартизованные остатки ☐ График подбора

Нормальная вероятность

☐ График нормальной вероятности

ОК

Отмена


Справка


Входной интервал Y –
диапазон, содержащий
данные результативного
признака

Использование инструмента РЕГРЕССИЯ

Регрессия

Входные данные

Входной интервал Y: 

Входной интервал X: 

☐ Метки

☐ Константа - ноль

☐ Уровень надежности: %

Остатки

☐ Остатки

☐ Стандартизованные остатки

☐ График остатков

☐ График подбора

Нормальная вероятность

☐ График нормальной вероятности

OK

Отмена

Справка

Входной интервал X –
диапазон, содержащий
данные фактора

Использование инструмента РЕГРЕССИЯ

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

☒ Метки

☐ Константа - ноль

☐ Уровень надежности: %

Параметры вывода

☐ Выходной интервал:

☒ Новый рабочий лист:

☐ График остатков

☐ График подбора

ОК

Отмена

Справка

Метки:

флажок, который указывает, содержит ли первая строка названия столбцов или нет

Использование инструмента РЕГРЕССИЯ

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

☐ Метки

☐ Константа - ноль

☐ Уровень надежности: %

Параметры вывода

☐ Выходной интервал:

☒ Новый рабочий лист:

☐ Новая рабочая книга

Остатки

☐ График остатков

☐ График подбора

OK

Отмена

Справка

Новый рабочий лист:
результаты можно вывести на
НОВЫЙ ЛИСТ

Использование инструмента РЕГРЕССИЯ

	A	B	C
1	По территориям		
2			
3	Номер региона	Среднедушевой минимум трудоспособного населения	
4	1		
5	2		
6	3		
7	4		
8	5		
9	6		
10	7		
11	8		
12	9		
13	10		
14	11		
15	12		
16			

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

☒ Метки

☐ Константа - ноль

☐ Уровень надежности: %

Параметры вывода

☐ Выходной интервал:

☒ Новый рабочий лист:

☐ Новая рабочая книга

Остатки

☐ Остатки

☒ График остатков

☐ Стандартизованные остатки

☐ График подбора

Нормальная вероятность

☐ График нормальной вероятности

OK

Отмена

Справка

Использование инструмента РЕГРЕССИЯ

The screenshot displays the Microsoft Excel interface with the following data:

Microsoft Excel - Пример 1 (откр лекция)

Введите во

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

100%

А B C D E F G H

1 ВЫВОД ИТОГОВ

2

3 Регрессионная статистика

4 Множественный R 0,721025214

5 R-квадрат 0,519877359

6 Нормированный R-квадрат 0,471865095

7 Стандартная ошибка 12,5495908

8 Наблюдения 12

9

10 Дисперсионный анализ

11

12 Регрессия 1 1705,327706 1705,327706 10,82801173 0,008141843

13 Остаток 10 1574,922294 157,4922294

14 Итого 11 3280,25

15

16

17 Коэффициенты Стандартная ошибка t-статистика P-Значение Нижние 95% Верхние 95%

18 Y-пересечение 76,9764852 24,21156138 3,179327594 0,009830668 23,02976485 130,923206 23,029765

19 Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., х 0,920430553 0,279715587 3,290594434 0,008141843 0,297185389 1,54367572 0,2971854

Использование инструмента РЕГРЕССИЯ

<i>Регрессионная статистика</i>	
Множественный R	0,721025214
R-квадрат	0,519877359
Нормированный R-квадрат	0,471865095
Стандартная ошибка	12,5495908
Наблюдения	12

Линейный коэффициент
корреляции r_{xy}
0.721

Использование инструмента РЕГРЕССИЯ

<i>Регрессионная статистика</i>	
Множественный R	0,721025214
R-квадрат	0,519877359
Нормированный R-квадрат	0,471865095
Стандартная ошибка	12,5495908
Наблюдения	12

Коэффициент детерминации

R^2

0.5198

Использование инструмента РЕГРЕССИЯ

Дисперсионный анализ				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Регрессия	1	1705,327706	1705,327706	10,82801173
Остаток	10	1574,922294	157,4922294	
Итого	11	3280,25		

F-статистика 10.828

Использование инструмента РЕГРЕССИЯ

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>
Y-пересечение	76,9764852	24,21156138	3,179327594
Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., x	0,920430553	0,279715587	3,290594434

t-статистика

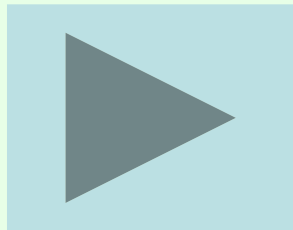
Параметр a:
76.976

Параметр b:
0.92

Уравнение регрессии: $y = 76.976 + 0.92x$

Демонстрация анализа с помощью инструмента РЕГРЕССИЯ

1. Ввод параметров и вывод результатов



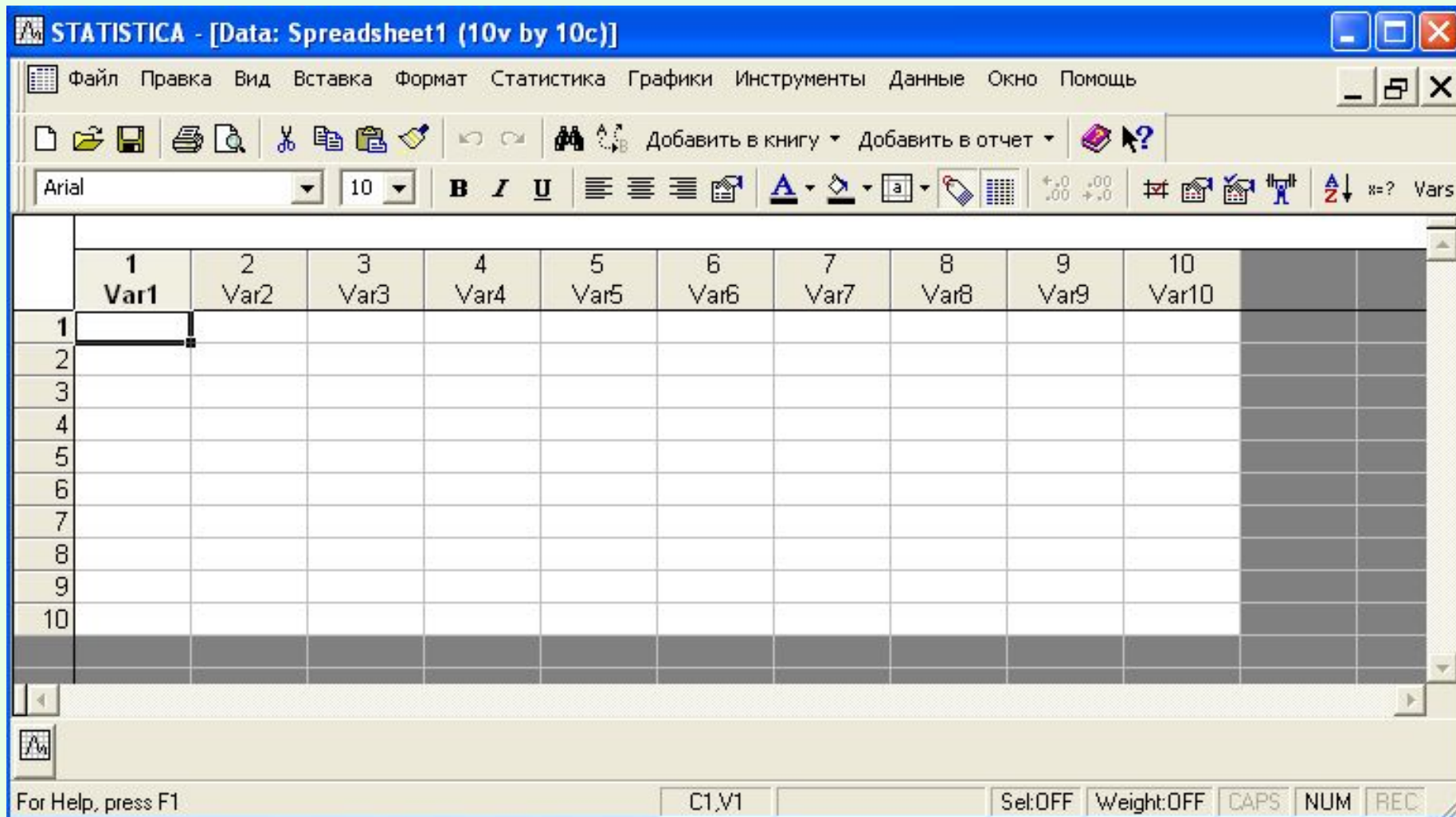
Парная линейная регрессия в MS Excel

3) Использование инструмента РЕГРЕССИЯ

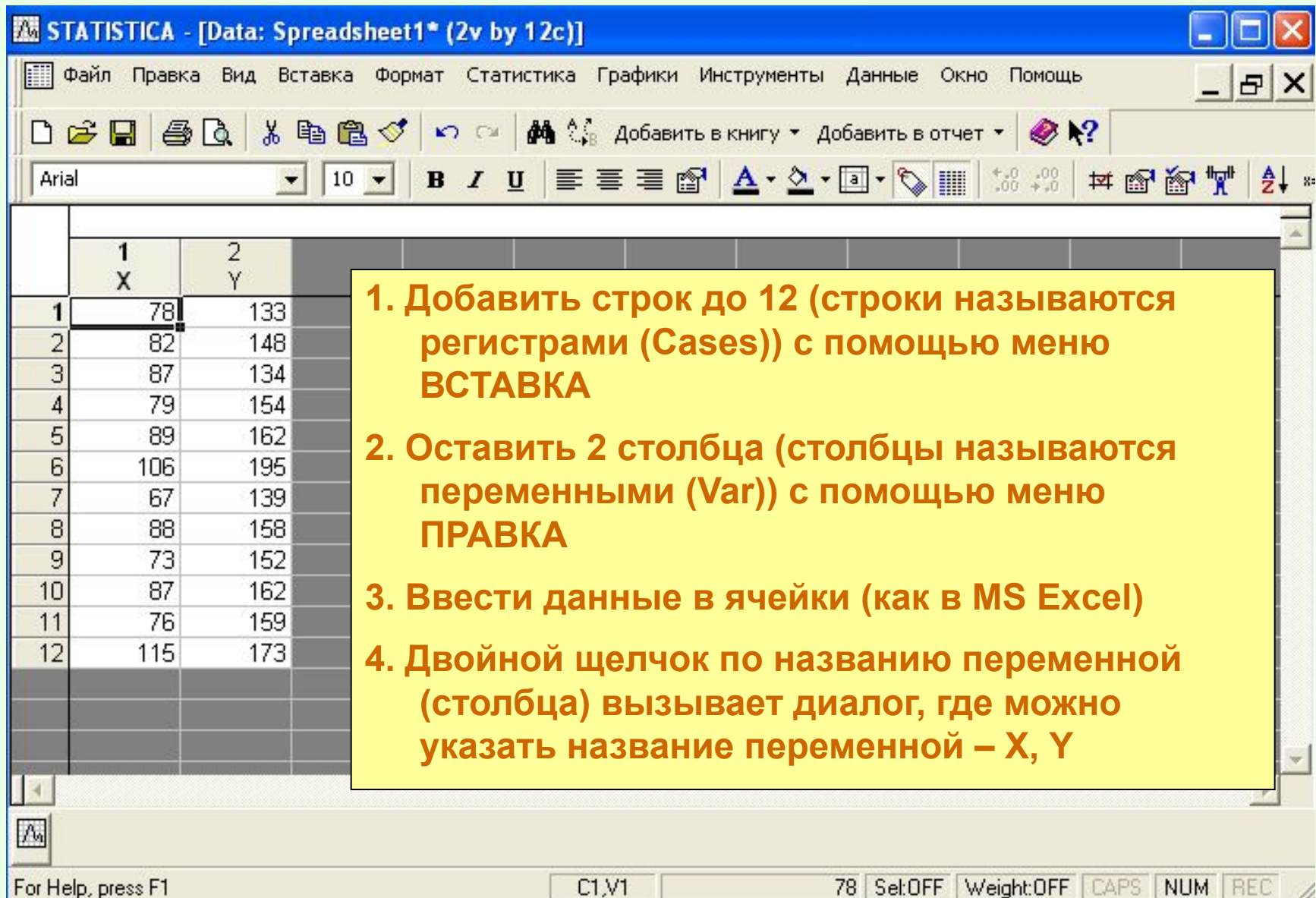
1. Ввести данные в 2 столбца
2. СЕРВИС / ПАКЕТ АНАЛИЗА / РЕГРЕССИЯ
3. Ввести параметры в окне РЕГРЕССИЯ,
нажать Ок

2. Использование STATISTICA для построения модели линейной регрессии

Окно программы STATISTICA



Ввод данных для анализа



STATISTICA - [Data: Spreadsheet1* (2v by 12c)]

Файл Правка Вид Вставка Формат Статистика Графики Инструменты Данные Окно Помощь

Добавить в книгу Добавить в отчет

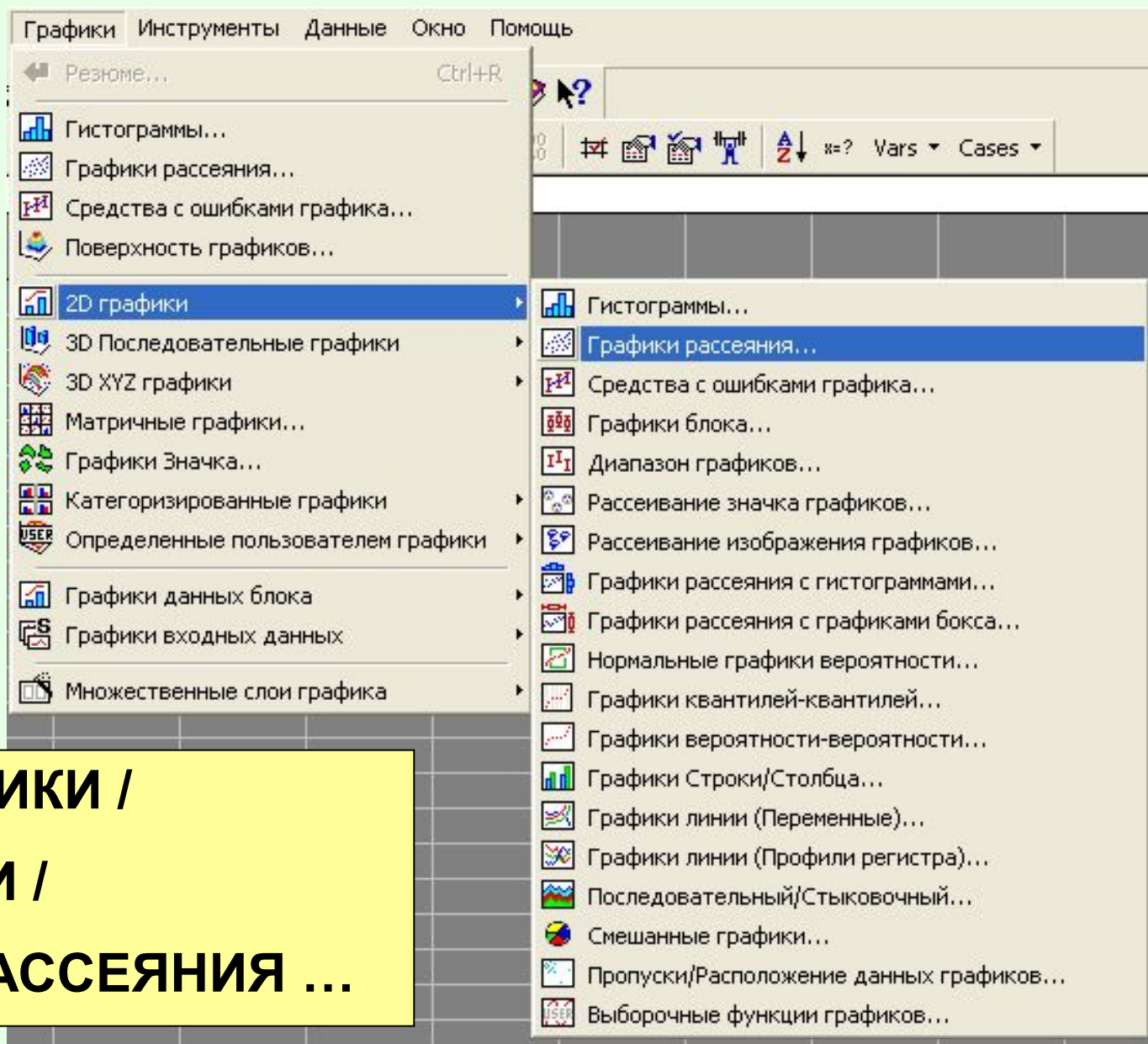
Arial 10 **B** *I* U [Text Alignment Icons] [Number Format Icons]

	1 X	2 Y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

1. Добавить строк до 12 (строки называются регистрами (Cases)) с помощью меню **ВСТАВКА**
2. Оставить 2 столбца (столбцы называются переменными (Var)) с помощью меню **ПРАВКА**
3. Ввести данные в ячейки (как в MS Excel)
4. Двойной щелчок по названию переменной (столбца) вызывает диалог, где можно указать название переменной – X, Y

For Help, press F1 C1.V1 78 Sel:OFF Weight:OFF CAPS NUM REC

Построение графика рассеяния переменных

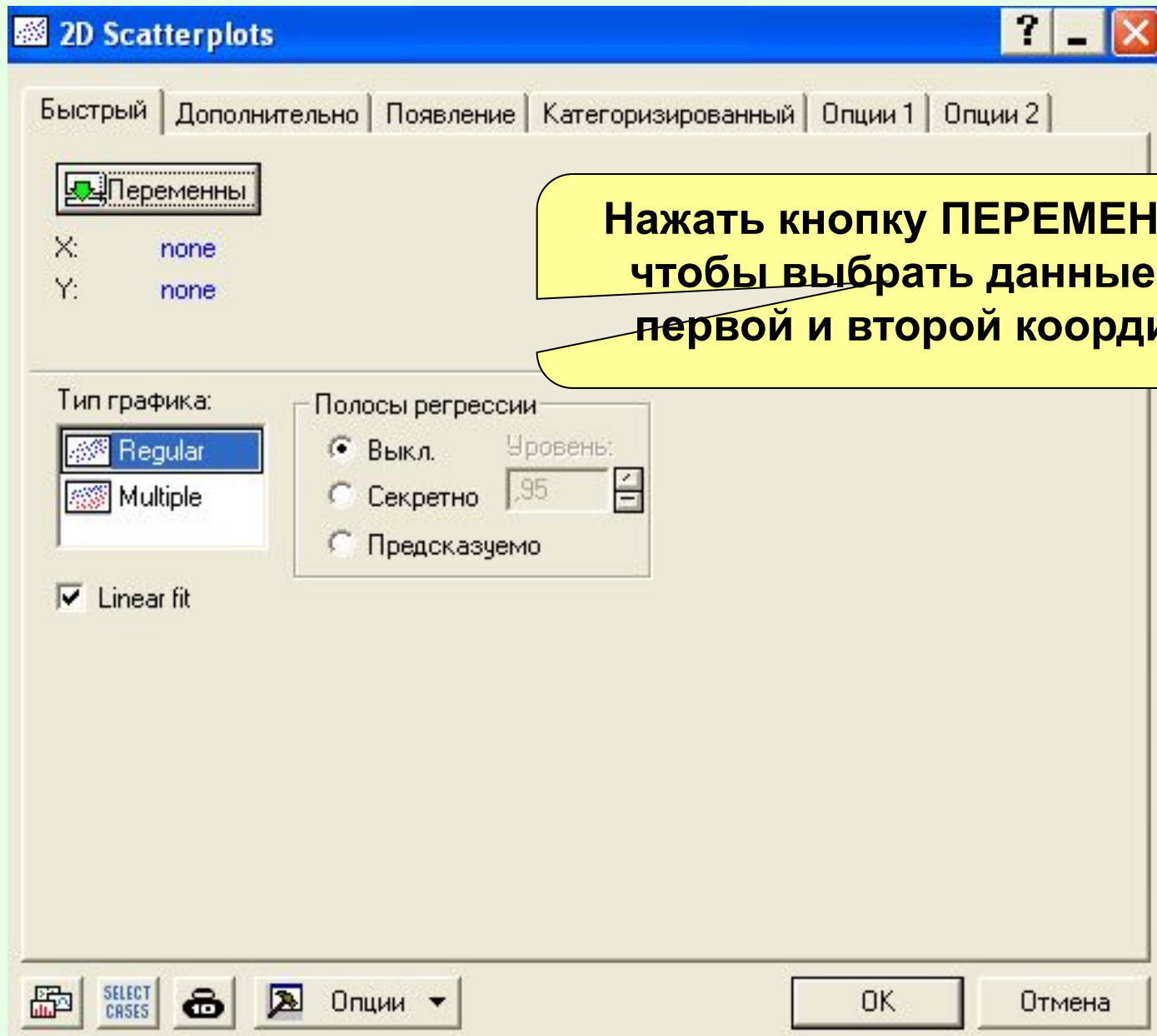


Меню ГРАФИКИ /

2D ГРАФИКИ /

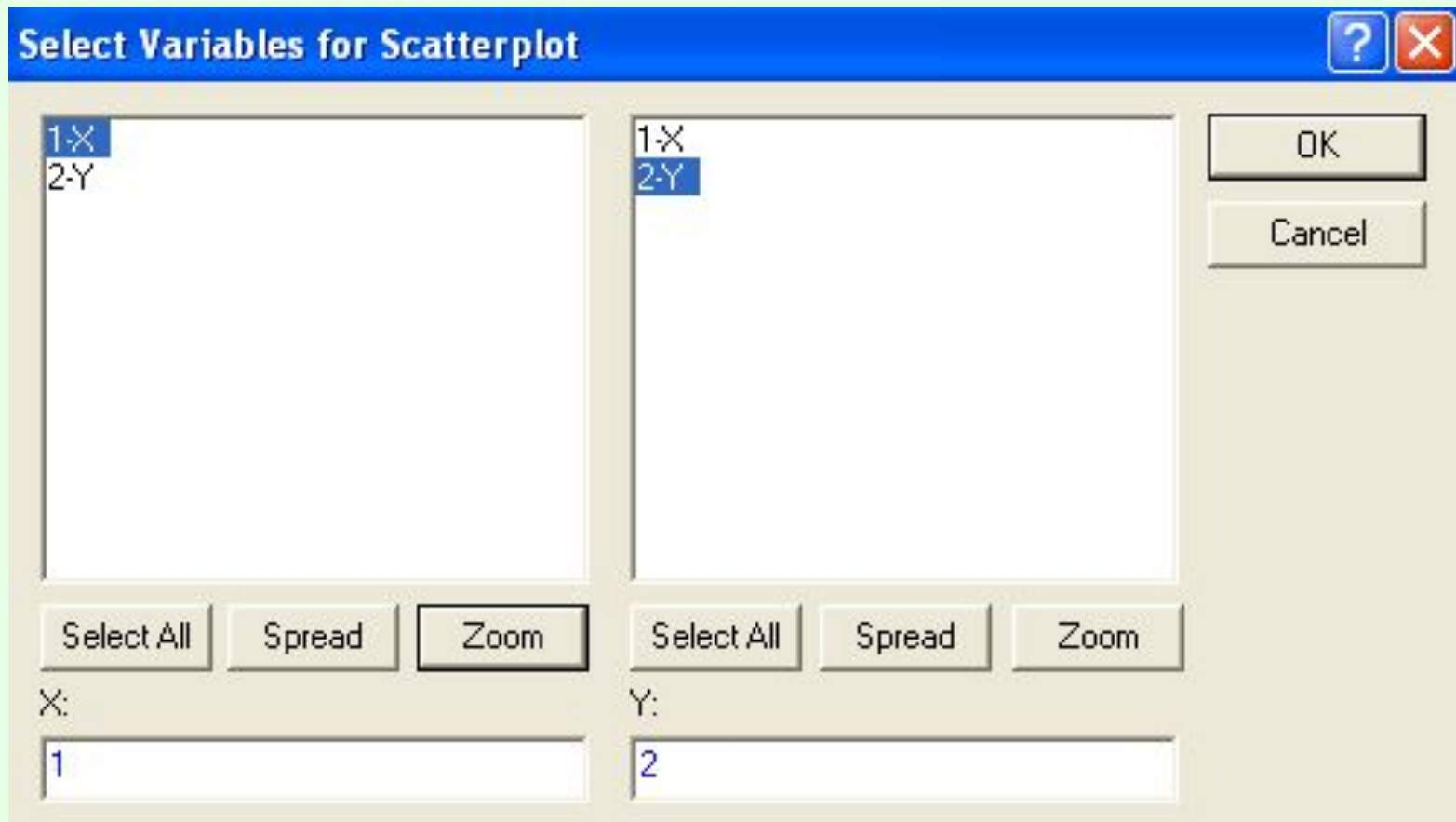
ГРАФИКИ РАССЕЯНИЯ ...

Построение графика рассеяния переменных



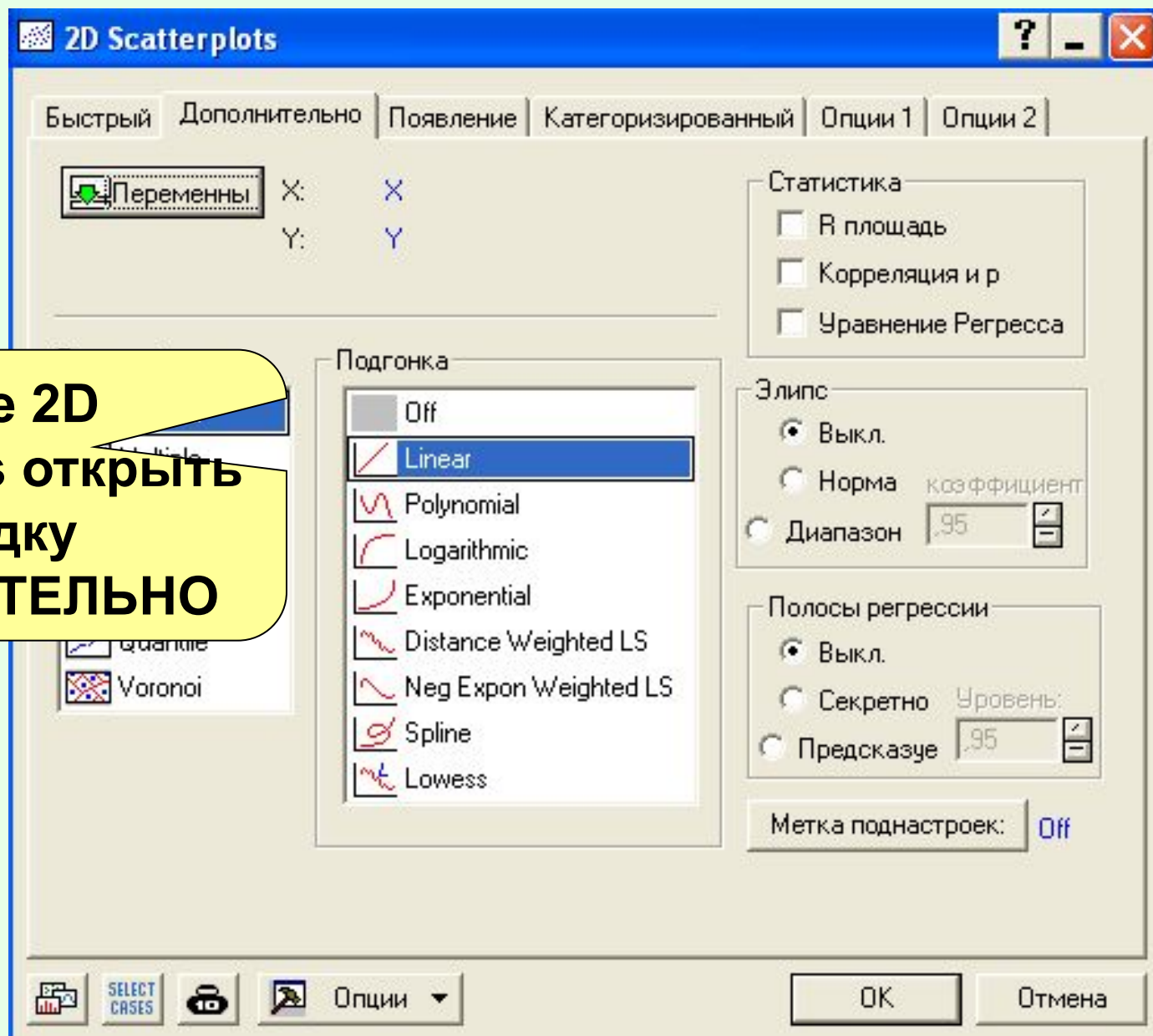
Построение графика рассеяния переменных

- В появившемся окне слева выбрать (выполнить щелчок) по переменной 1-X, а в правом окне – по переменной 2-Y.
- Нажать Ок



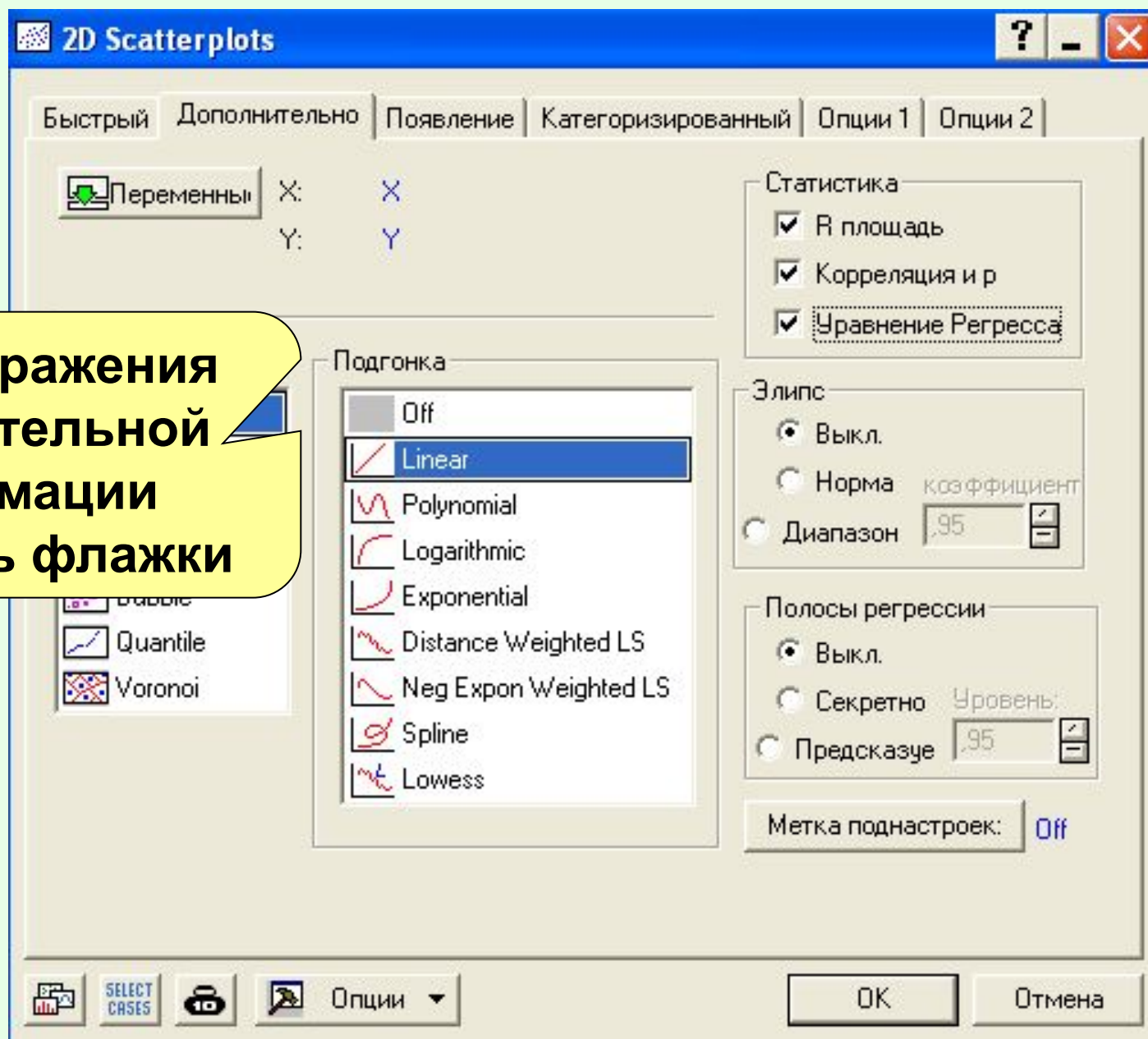
Построение графика рассеяния переменных

В окне 2D Scatterplots открыть вкладку ДОПОЛНИТЕЛЬНО



Построение графика рассеяния переменных

Для отображения
дополнительной
информации
поставить флажки



Построение графика рассеяния переменных

Scatterplot (Spreadsheet1 2v*12c)

$$Y = 76,9765 + 0,9204 * x$$

- Нажать Ок и появится диаграмма
- Отображается уравнение, коэффициент корреляции и коэффициент детерминации



X:Y: $r^2 = 0,5199$; $r = 0,7210$; $p = 0,0081$; $y = 76,9764852 + 0,920430553 * x$

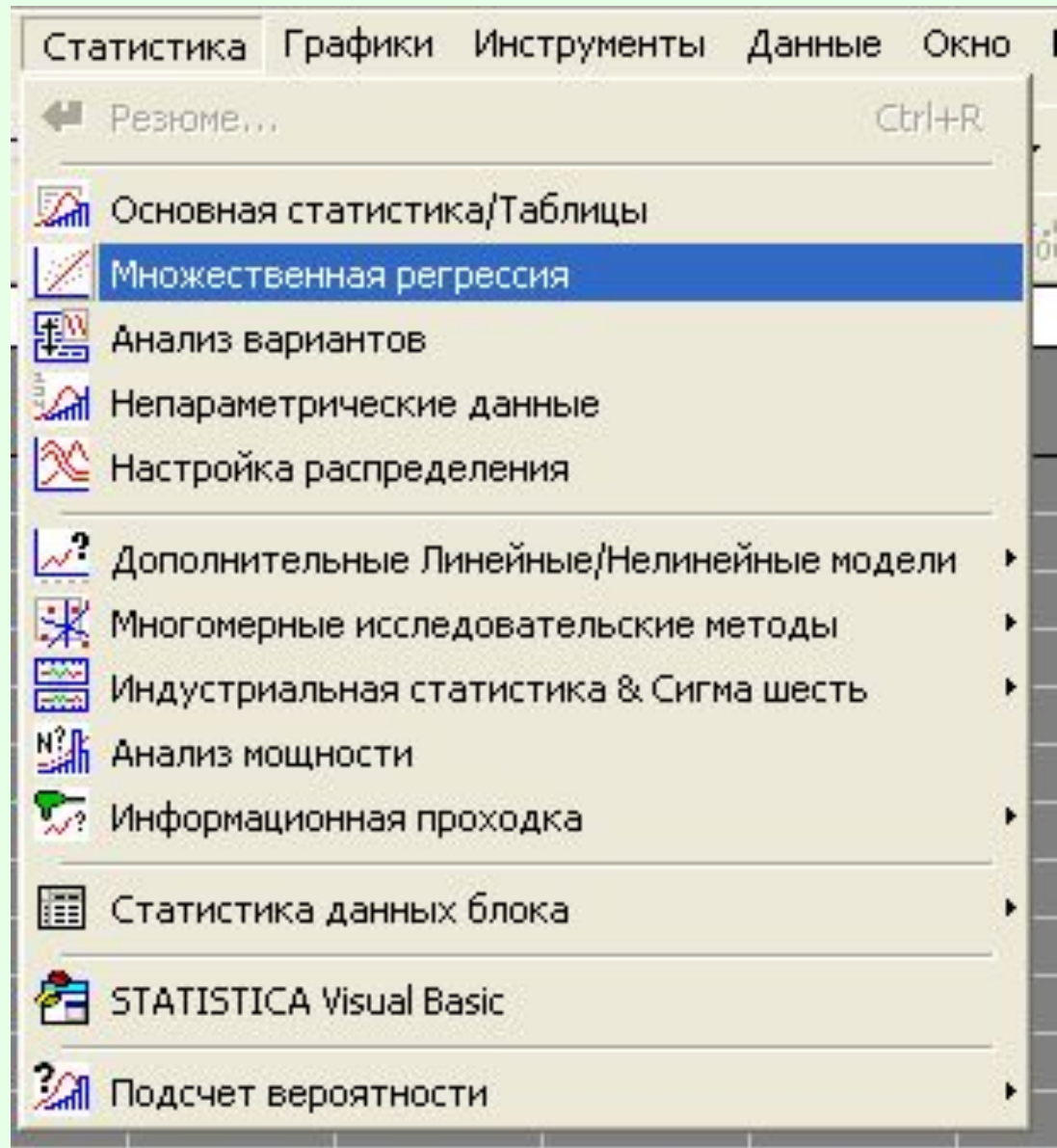
Линейная регрессия в STATISTICA

1) Построение диаграммы рассеяния

1. Ввести данные
2. ГРАФИКИ / 2D ГРАФИКИ / ГРАФИКИ РАССЕЯНИЯ ...
3. Выбрать переменные на вкладке БЫСТРЫЙ
4. На вкладке ДОПОЛНИТЕЛЬНО поставить флажки на отображении статистики
5. Нажать Ок

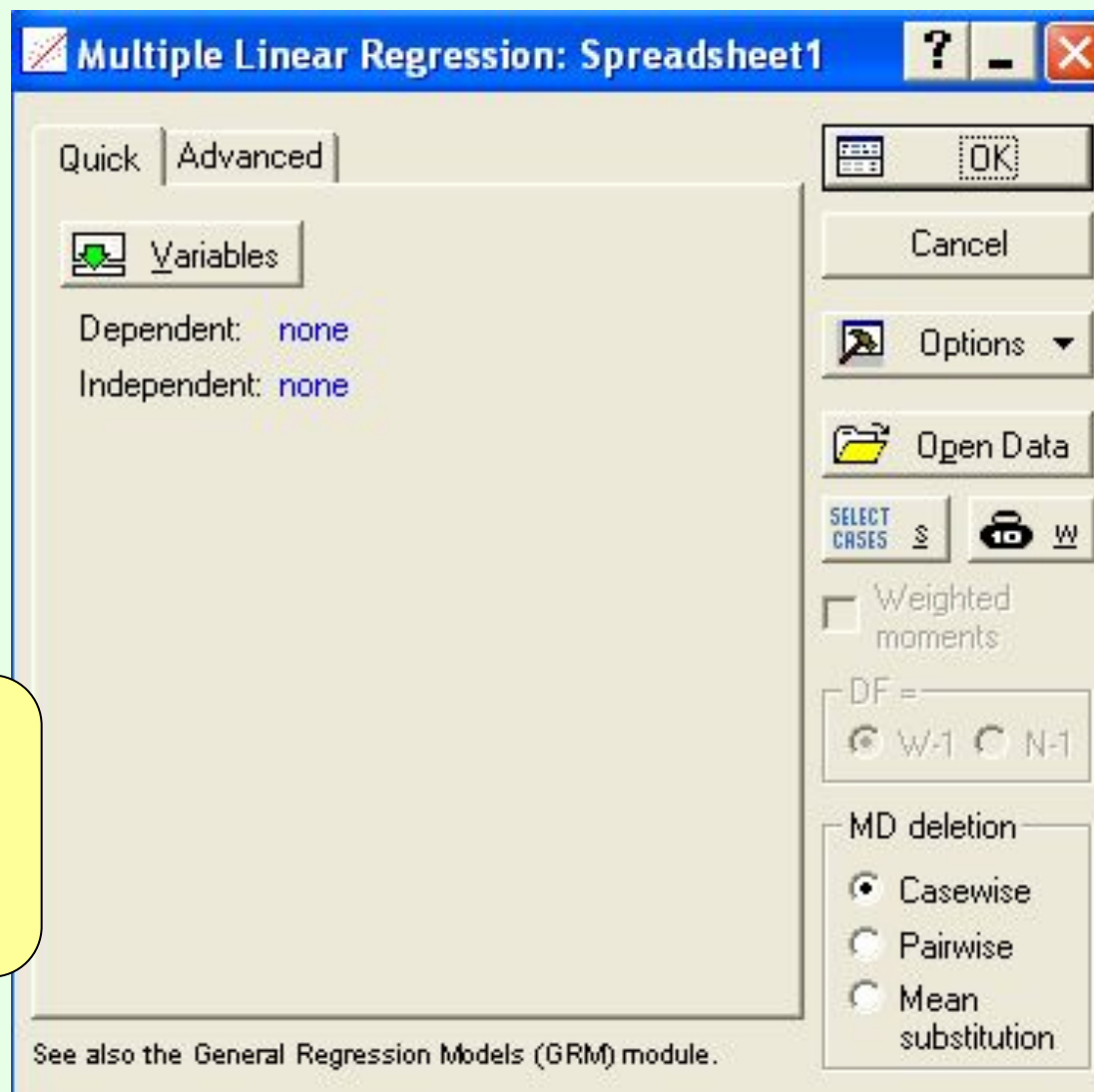
Регрессионный анализ

- Меню
СТАТИСТИКА
- Множественная
регрессия



Регрессионный анализ

**Нажать кнопку
Variables
(Переменные)**



Регрессионный анализ

Select dependent and independent variable lists: ? X

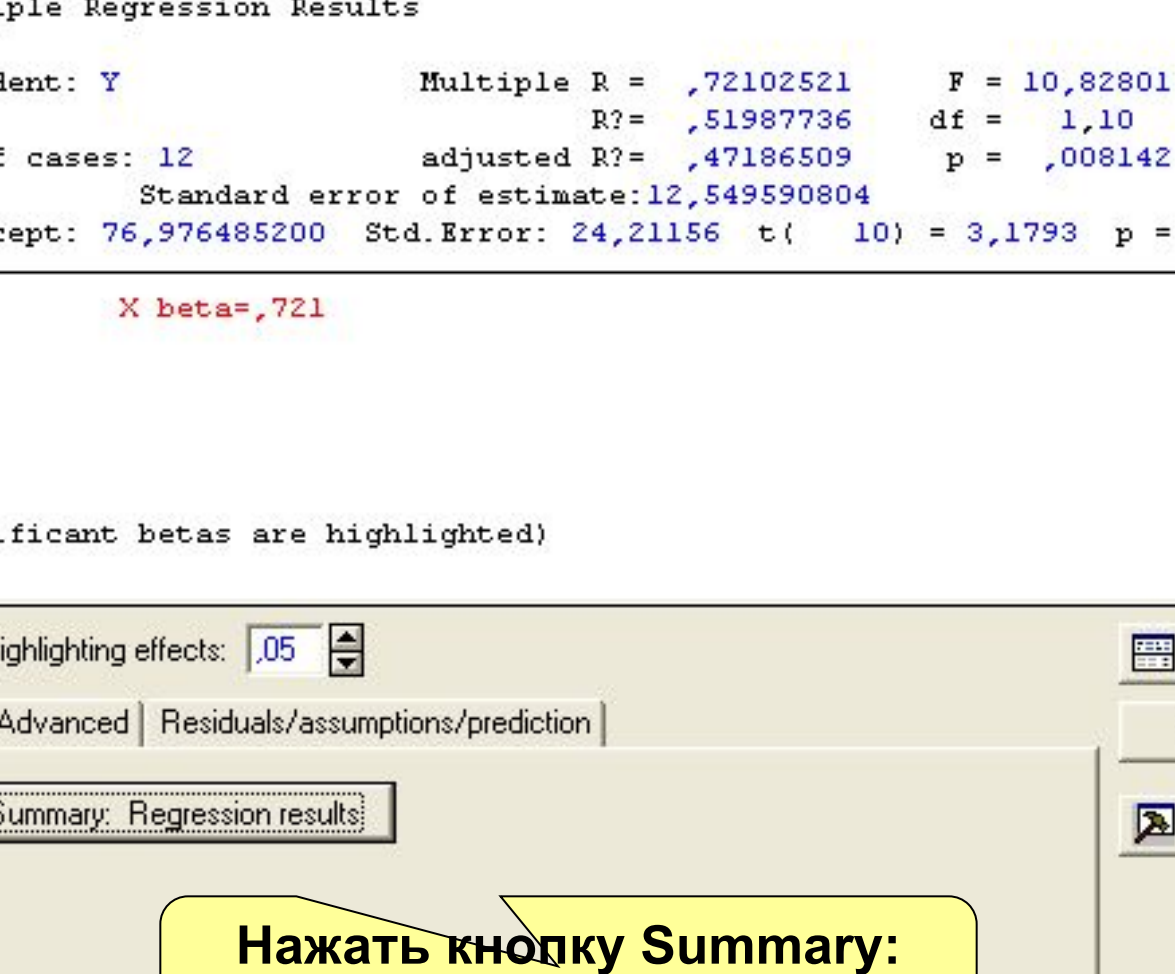
Dependent var. (or list for batch):	Independent variable list:
<div>1-X 2-Y</div> <div>Select All Spread Zoom</div> <div>2</div>	<div>1-X 2-Y</div> <div>Select All Spread Zoom</div> <div>1</div>

OK
Cancel

**Слева указать переменную
Y (Dependent – зависимая
переменная)**

**Справа указать переменную
X (Independent –
независимая переменная)**

Регрессионный анализ



Multiple Regression Results: Spreadsheet1

Multiple Regression Results

Dependent: Y Multiple R = ,72102521 F = 10,82801
R² = ,51987736 df = 1,10
No. of cases: 12 adjusted R² = ,47186509 p = ,008142
Standard error of estimate: 12,549590804
Intercept: 76,976485200 Std. Error: 24,21156 t(10) = 3,1793 p = ,0098

X beta = ,721

(significant betas are highlighted)

Alpha for highlighting effects: .05

Quick | Advanced | Residuals/assumptions/prediction

Summary: Regression results

OK
Cancel
Options

**Нажать кнопку Summary:
Regression result**

Регрессионный анализ

STATISTICA - [Пример 1 (откр лекция)* - Regression Summary for Dependent Variable: Y (Spreadsheet1)]

File Правка Вид Вставка Формат Статистика Графики Инструменты Данные Workbook Окно Помощь

Добавить в книгу Добавить в отчет

Regression Summary for Dependent Variable: Y (Spreadsheet1)
R= ,72102521 R²= ,51987736 Adjusted R²= ,47186509
F(1,10)=10,828 p<,00814 Std. Error of estimate: 12,550

N=12	Beta	Std. Err. of Beta	B	Std. Err. of B	t(10)	p-level		
Intercept			76,97649	24,21156	3,179328	0,009831		
X	0,721025	0,219117	0,92043	0,27972	3,290594	0,008142		

Параметры регрессии

t-статистика

Линейная регрессия в STATISTICA

2) Регрессионный анализ

1. Ввести данные
2. СТАТИСТИКИ / МНОЖЕСТВЕННАЯ РЕГРЕССИЯ
3. Выбрать переменные (Dependent – зависимая; Independent – независимая)
4. Отобразить результаты: кнопка Summary: Regression result