Использование прикладного ПО для решения задач по теме «Линейная парная регрессия»

Цели лекции

1. Познакомить студентов с возможностями прикладного ПО (MS Excel, Statistica) для регрессионного анализа (на примере парной линейной регрессии)
2. Рассмотреть конкретные алгоритмы выполнения расчетов в указанных программах

Вопросы лекции

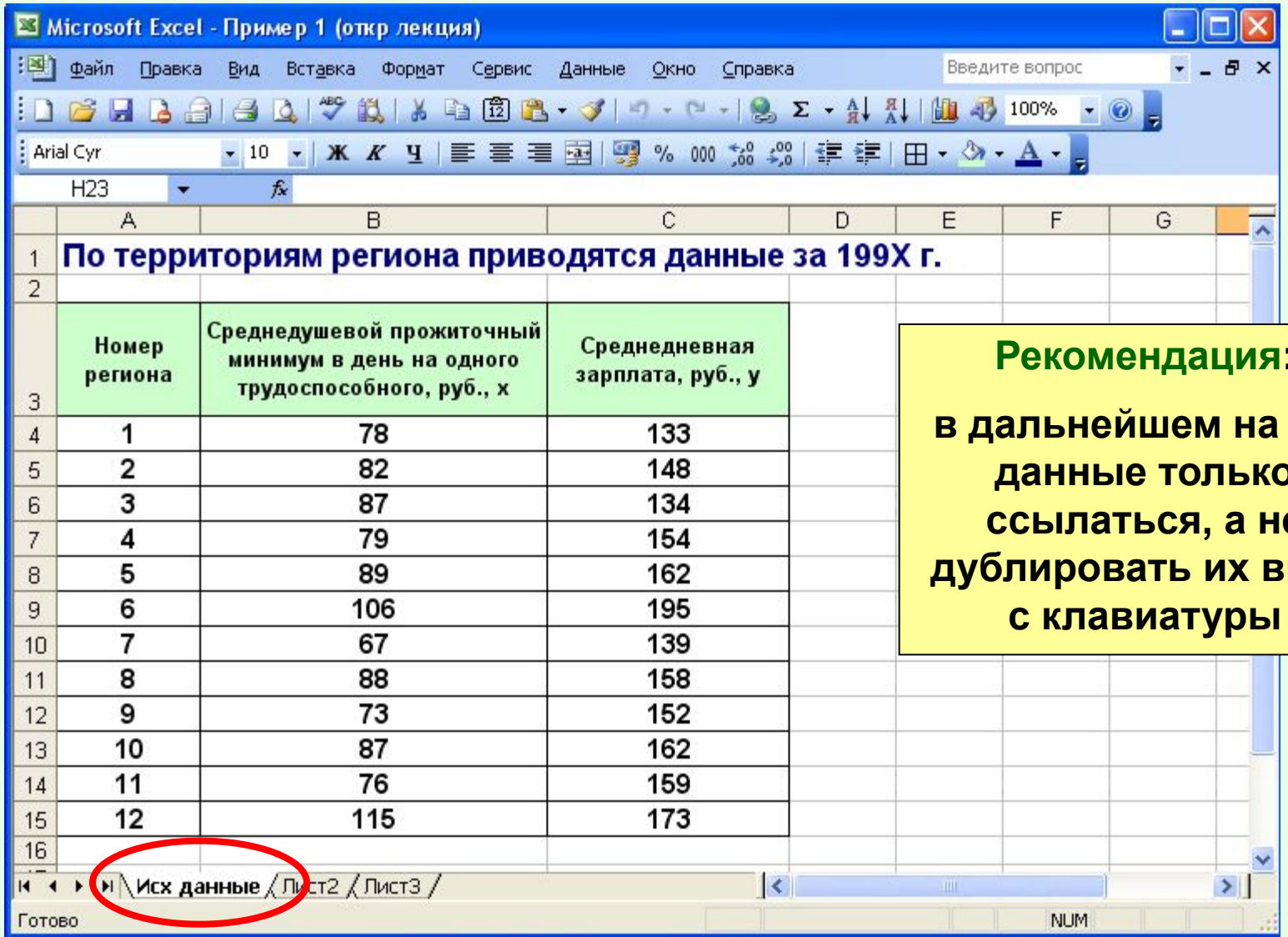
1. Использование электронных таблиц MS Excel для решения задач по теме «Линейная парная регрессия»
2. Использование программы STATISTICA для решения задач по теме «Линейная парная регрессия»

Литература:

1. Макарова Н.В., Трофимец В.Я. Статистика в Excel. - М.: Финансы и статистика, 2002. - 368 с.
2. Гельман В.Я. Решение математических задач средствами Excel. - СПб., 2003. - 236 с.
3. Боровиков В.П. Популярное введение в программу STATISTICA. - М., 2000. - 269 с.
4. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. - М.: Форум, 2008. - 464 с.
5. Елисеева И.И. Практикум по эконометрике: Учебное пособие. - М.: Финансы и статистика, 2005 - 192 с.

1. Использование MS Excel для решения задач по теме «Линейная парная регрессия»

На отдельном листе вводятся и оформляются исходные данные



Рекомендация:

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

Предварительный анализ

1. Построить корреляцио
   визуальный анализ

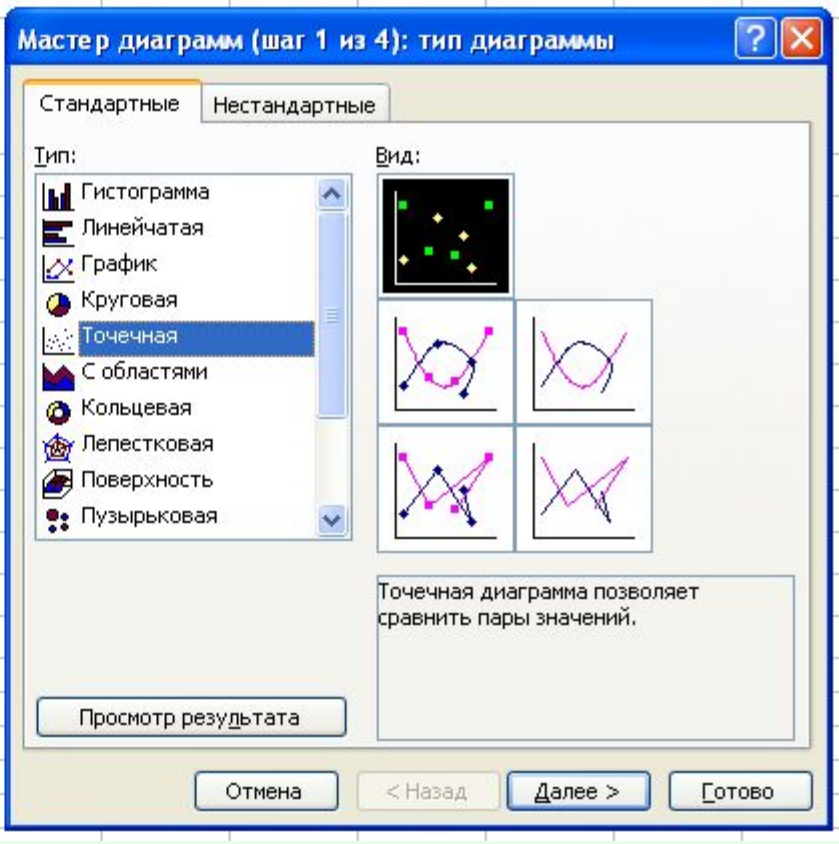
-Выделить диапазон
(2 столбца данных)

-Вызвать мастер диаграмм

-Выбрать точечную диаграмму

По территориям региона приводятся данные за 199X г.

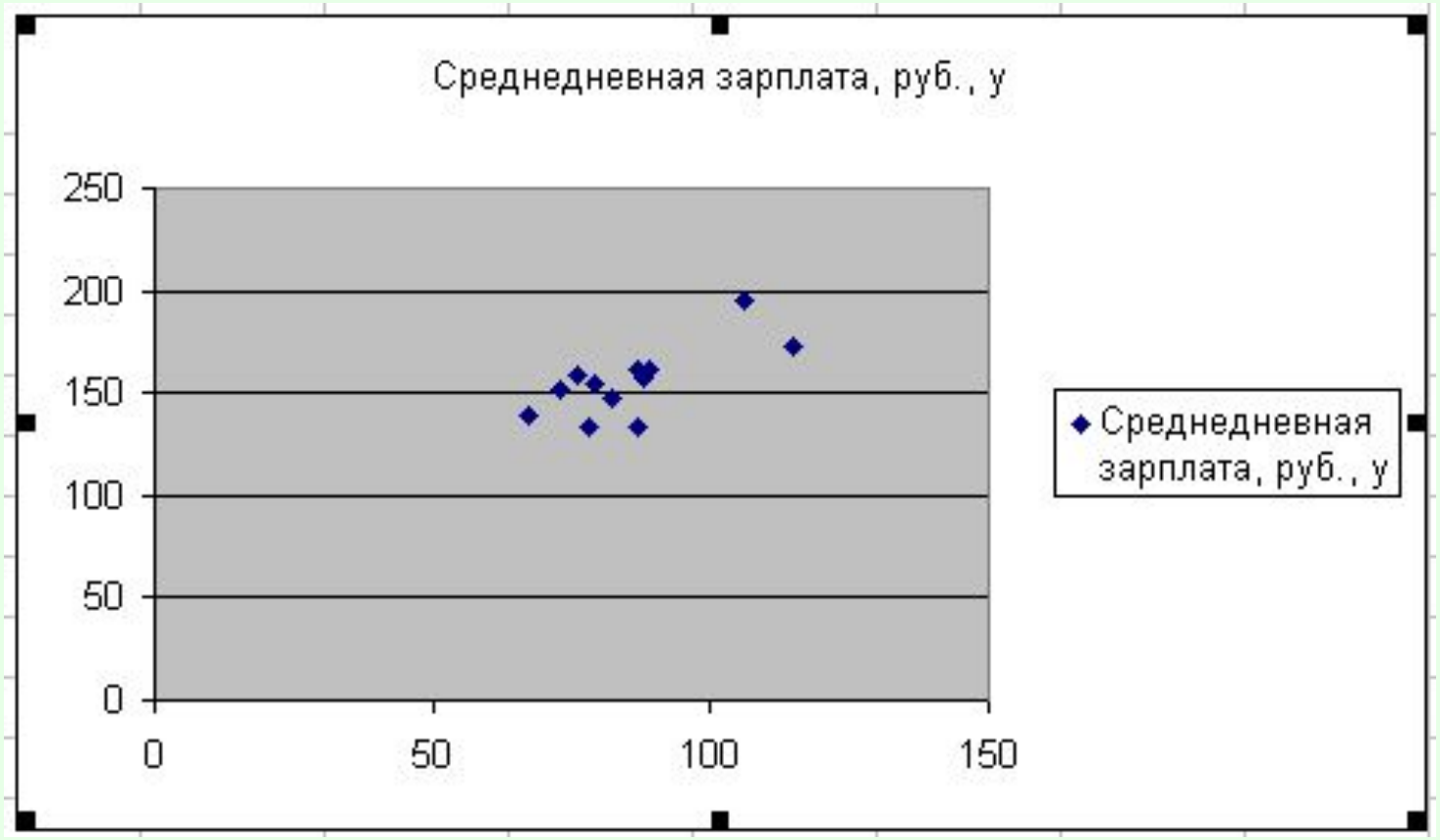
Номер региона	Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., x	Среднедневная зарплата, руб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173



Предварительный анализ

1. Построить корреляционное поле – визуальный анализ

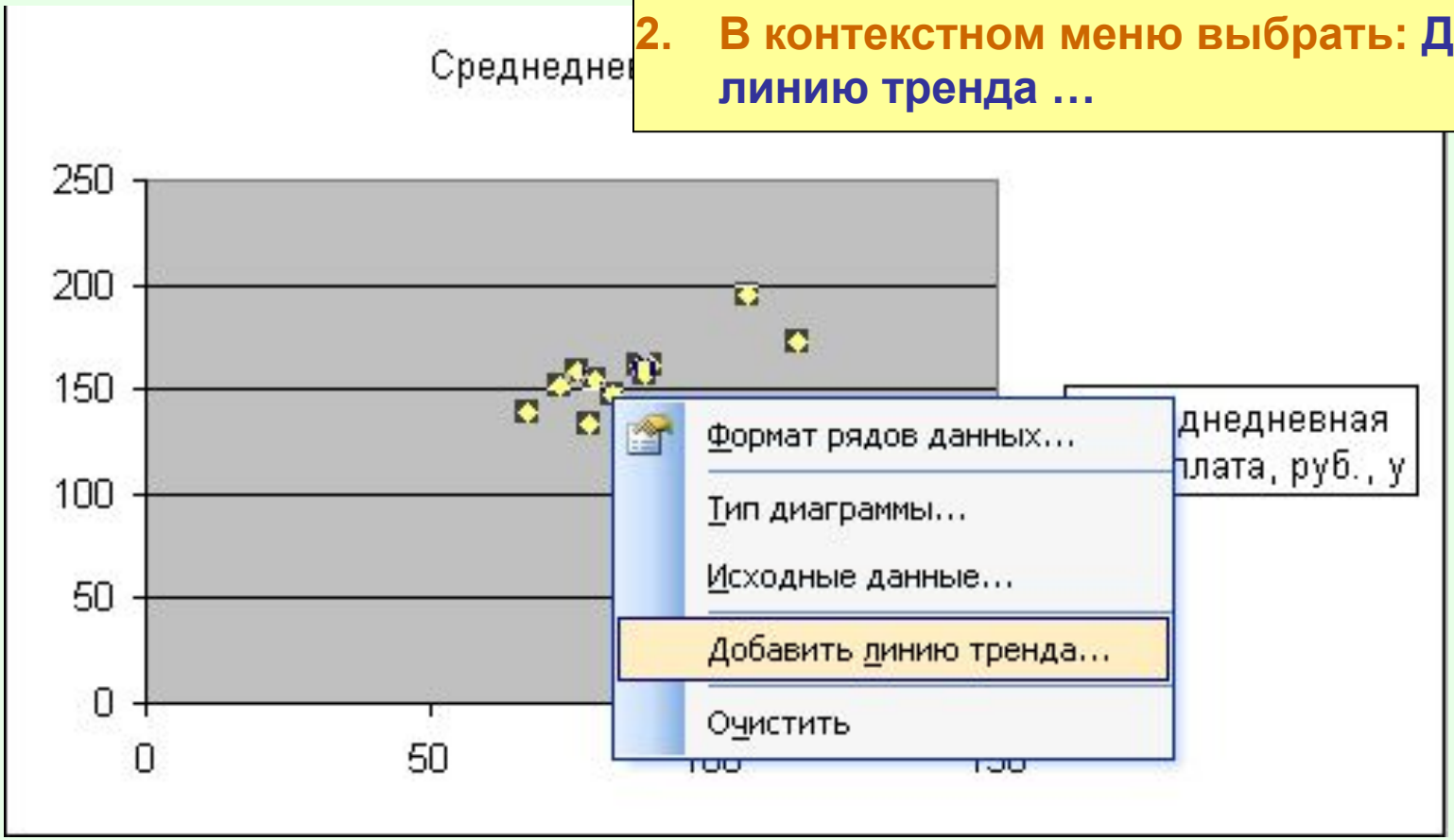
Нажать кнопку «Готово» и появится диаграмма – корреляционное поле



Предварительный анализ

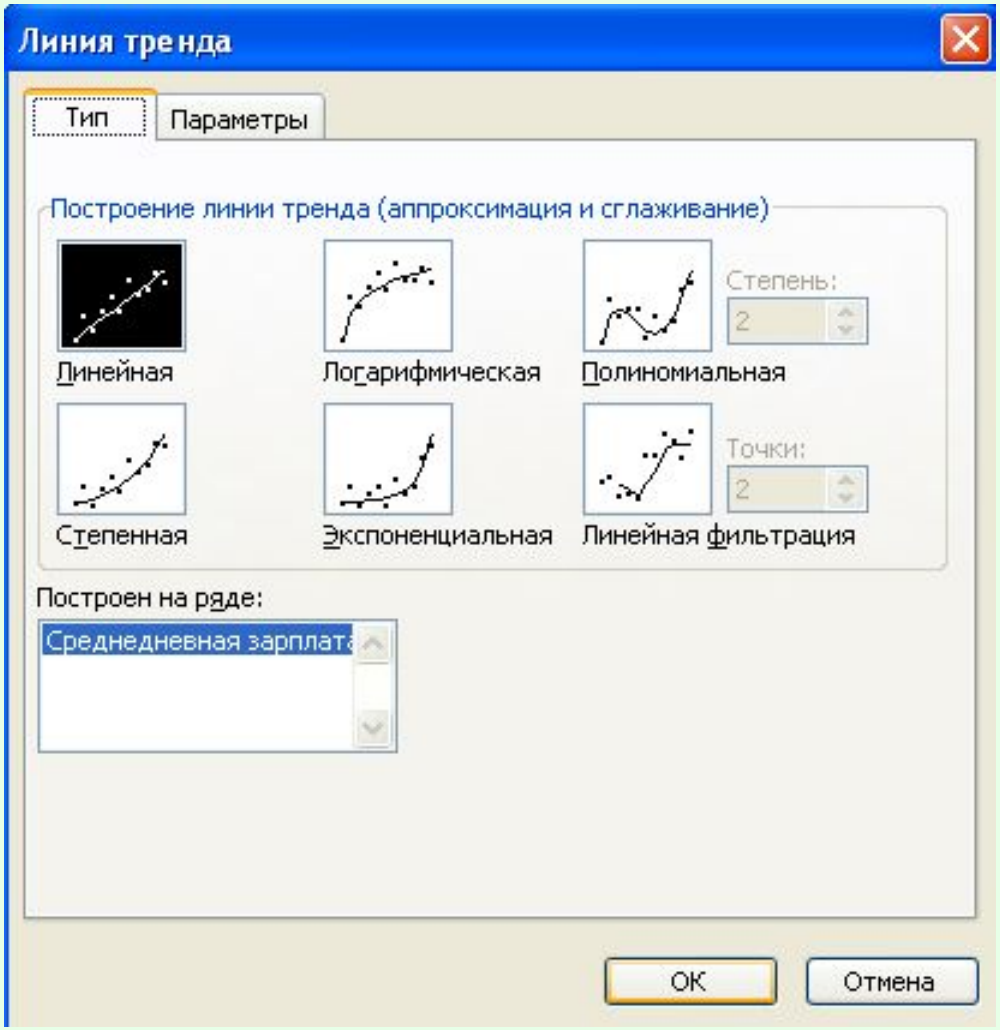
2. Добавить линию тренда, показать его уравнение и коэффициент детерминации

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



Предварительный анализ

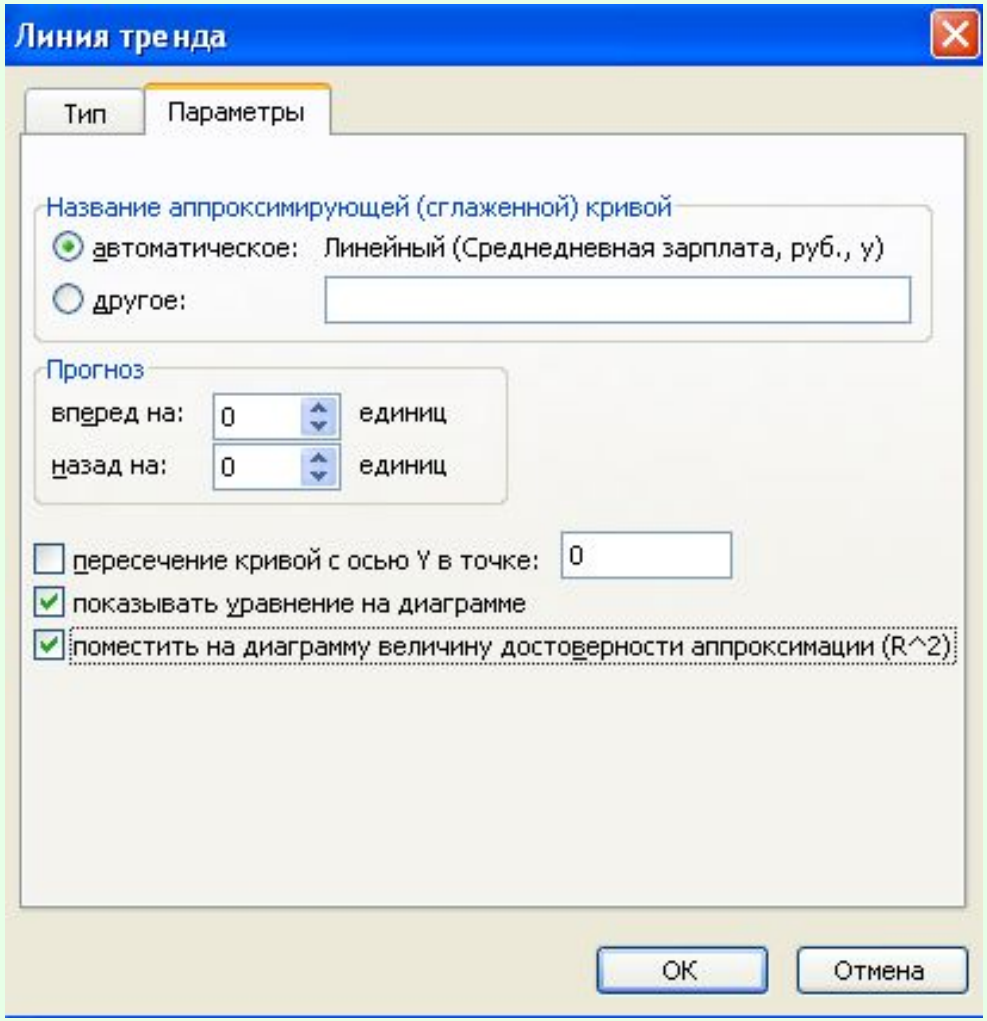
2. Добавить линию тренда, показать его уравнение и коэффициент детерминации



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

Предварительный анализ

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

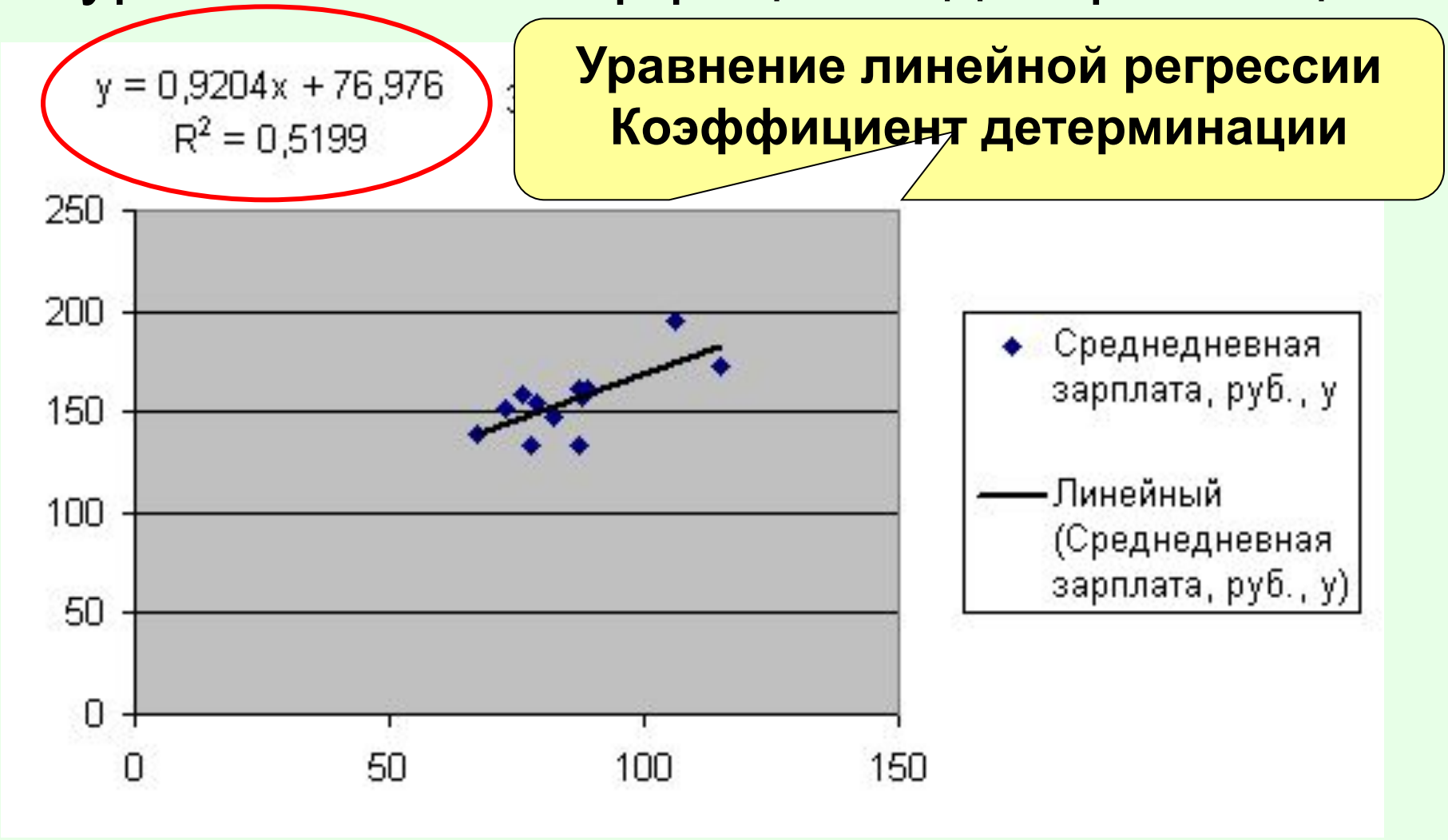


На вкладке «ПАРАМЕТРЫ» установить флажки:

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

Предварительный анализ

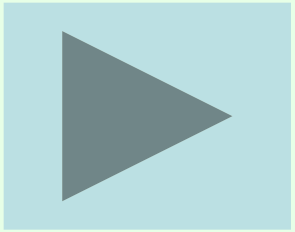
2. Добавить линию тренда, показать его уравнение и коэффициент детерминации



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

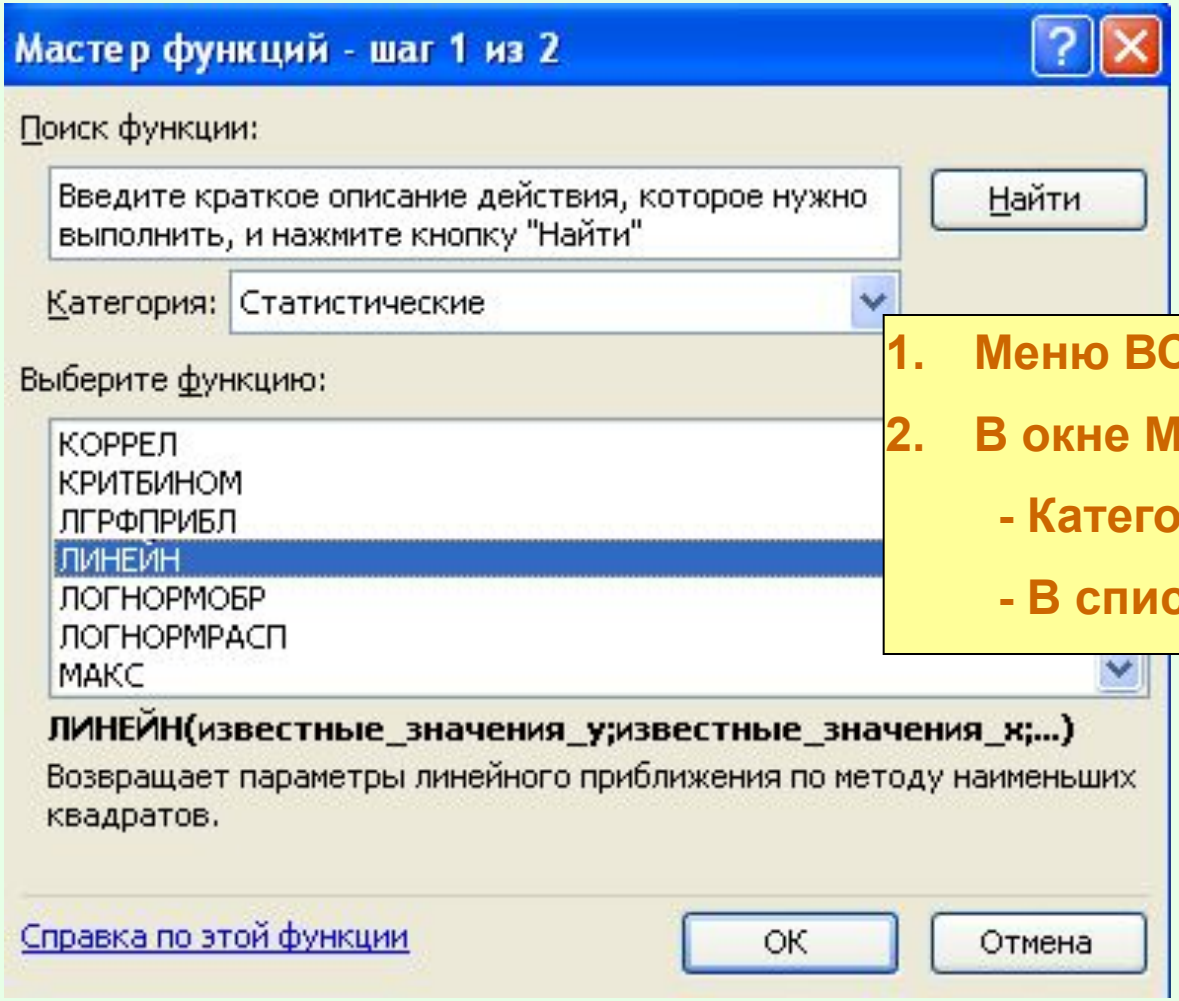
1) Использование диаграммы

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

Записать в тетрадь

Линейный регрессионный анализ

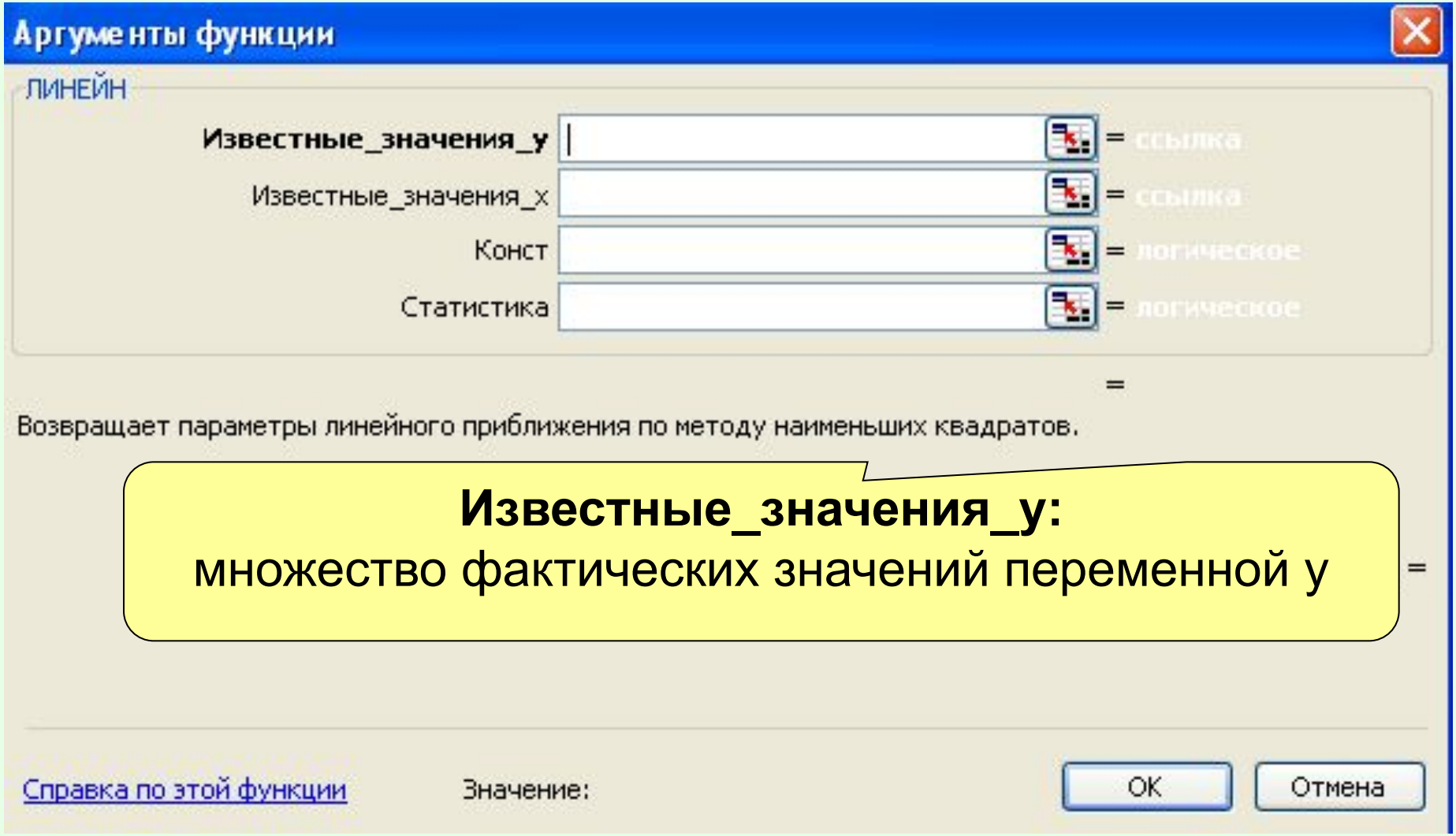
Используется встроенная статистическая функция ЛИНЕЙН, которая определяет параметры линейной регрессии



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

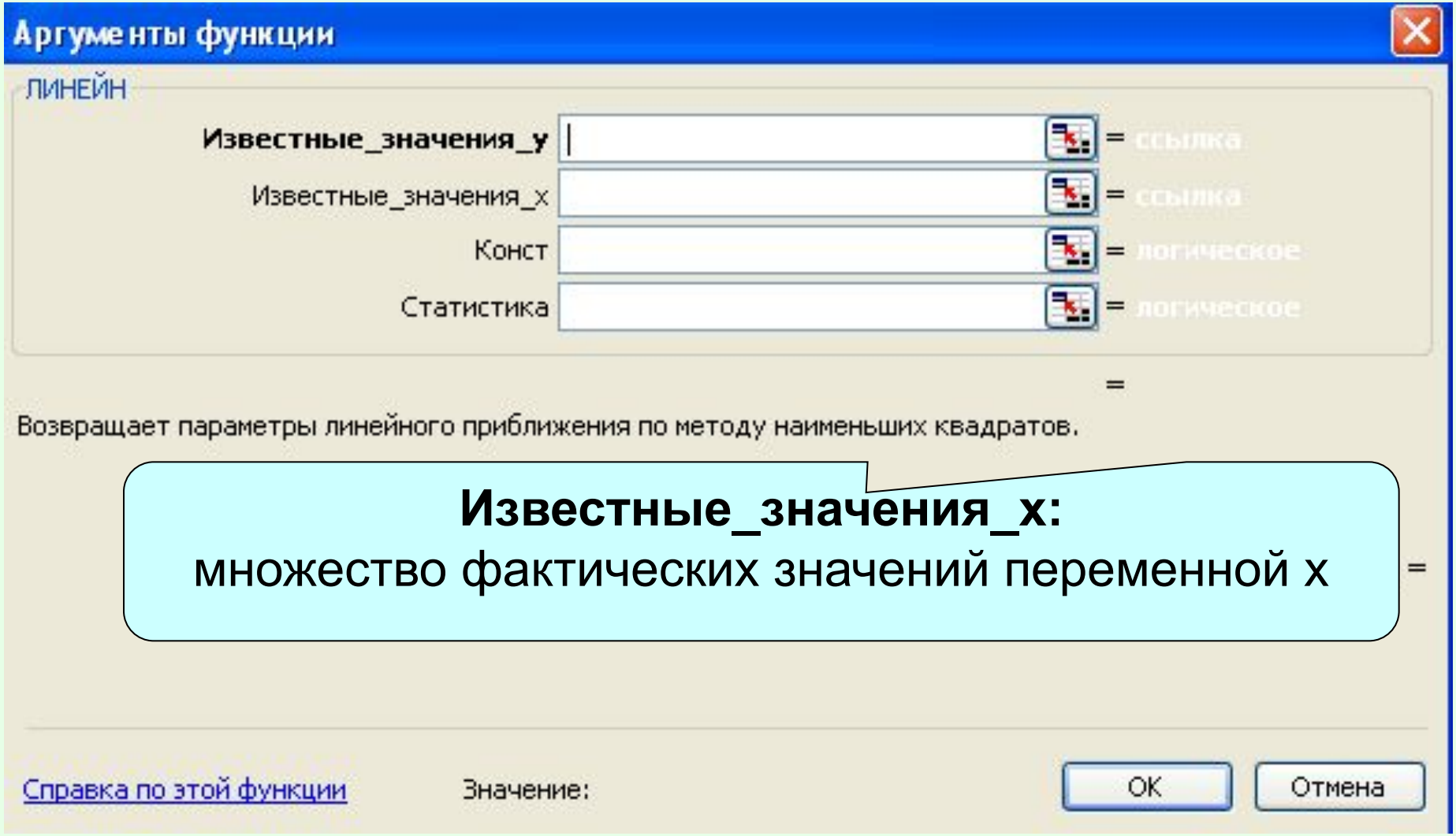
Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН



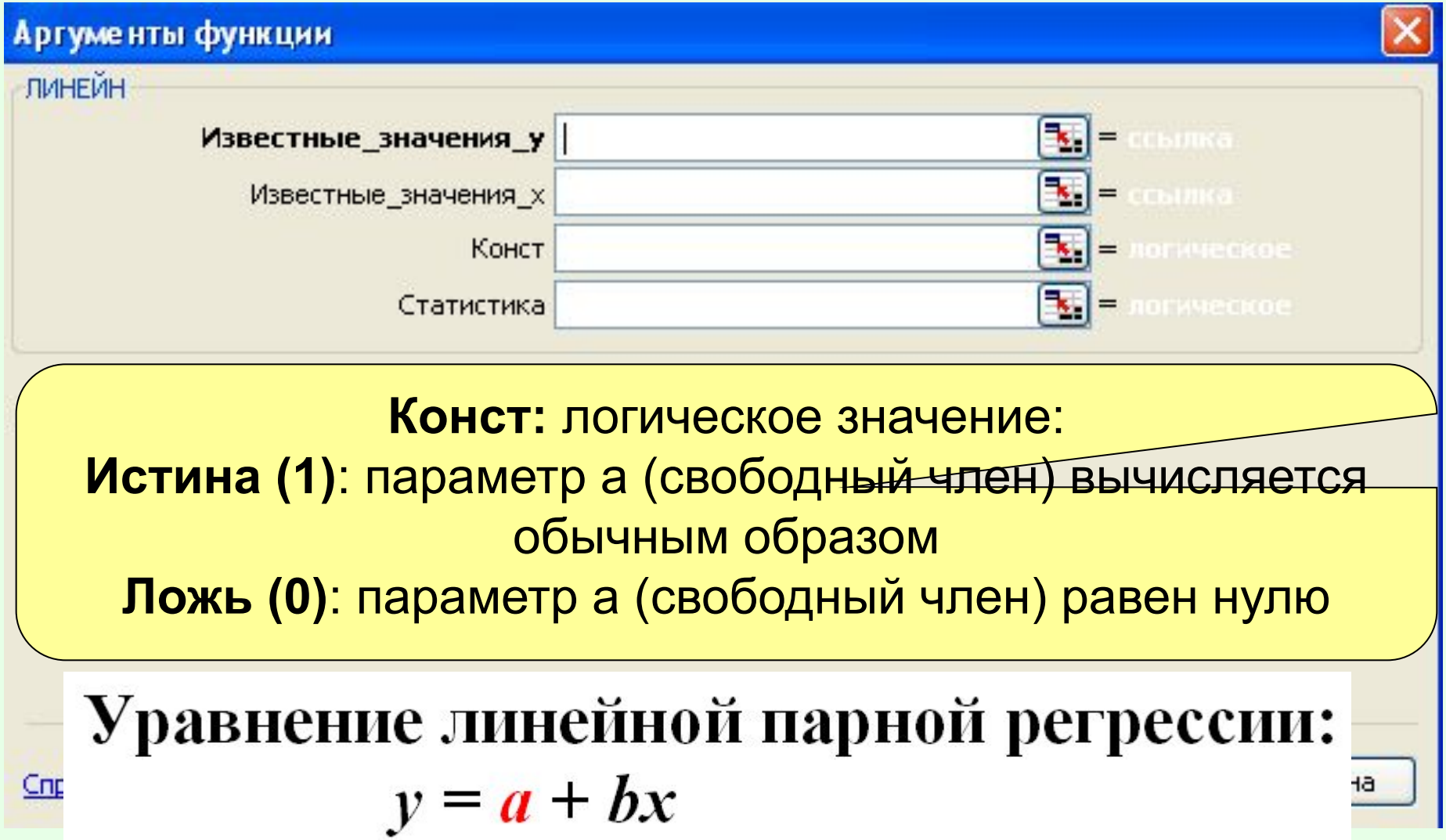
Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН



Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН



Конст: логическое значение:

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

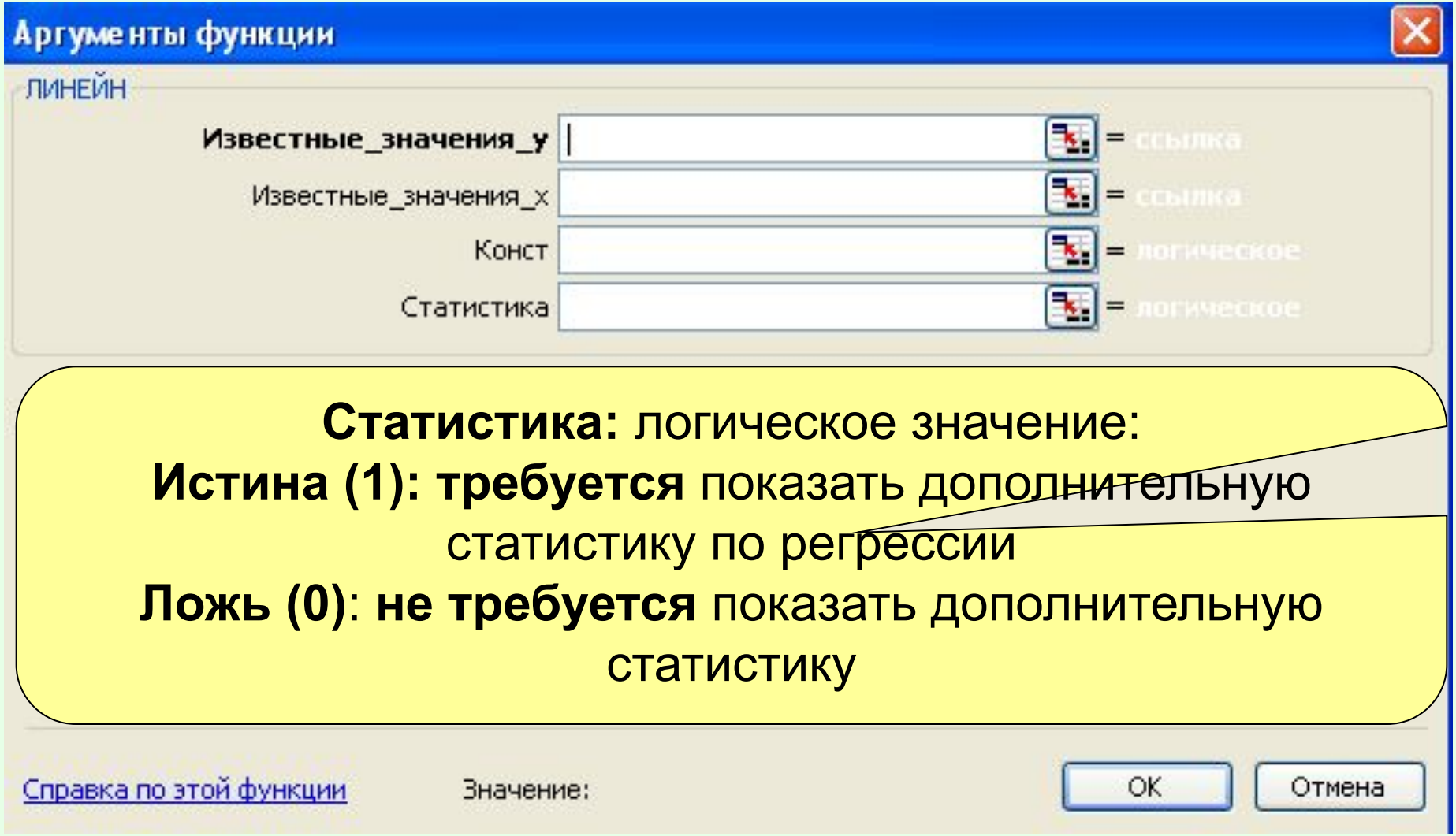
Ложь (0): параметр a (свободный член) равен нулю

Уравнение линейной парной регрессии:

$$y = a + bx$$

Линейный регрессионный анализ

Статистическая функция ЛИНЕЙН



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		Регрессионаая сумма квадратов (RSS)						Остаточная сумма квадратов (ESS)				

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

Уравнение линейной парной регрессии:

$$y = a + bx$$

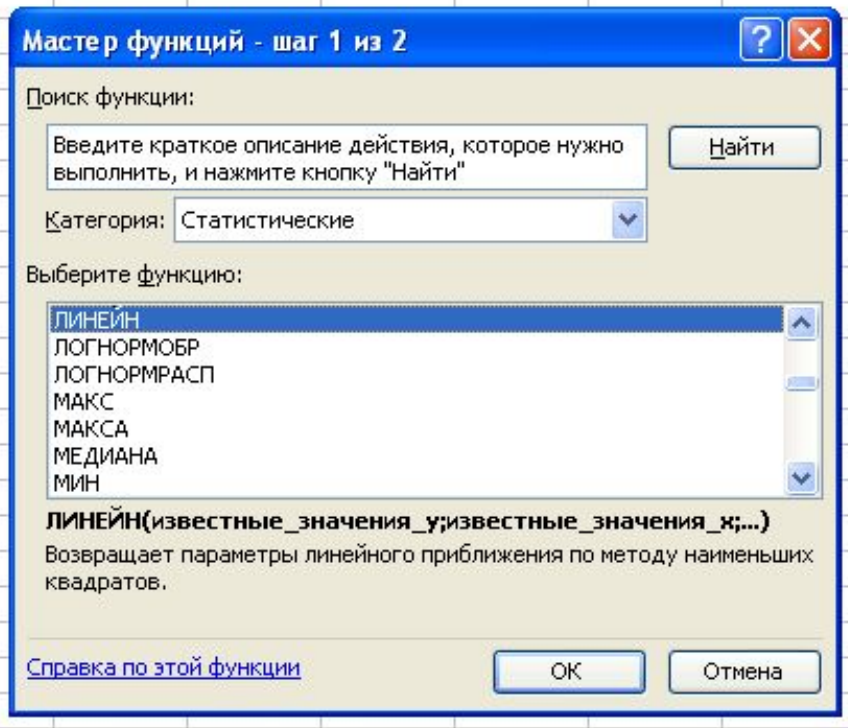
Линейный регрессионный анализ

Ход ввода параметров функции ЛИНЕЙН

Результаты линейного регрессионного анализа

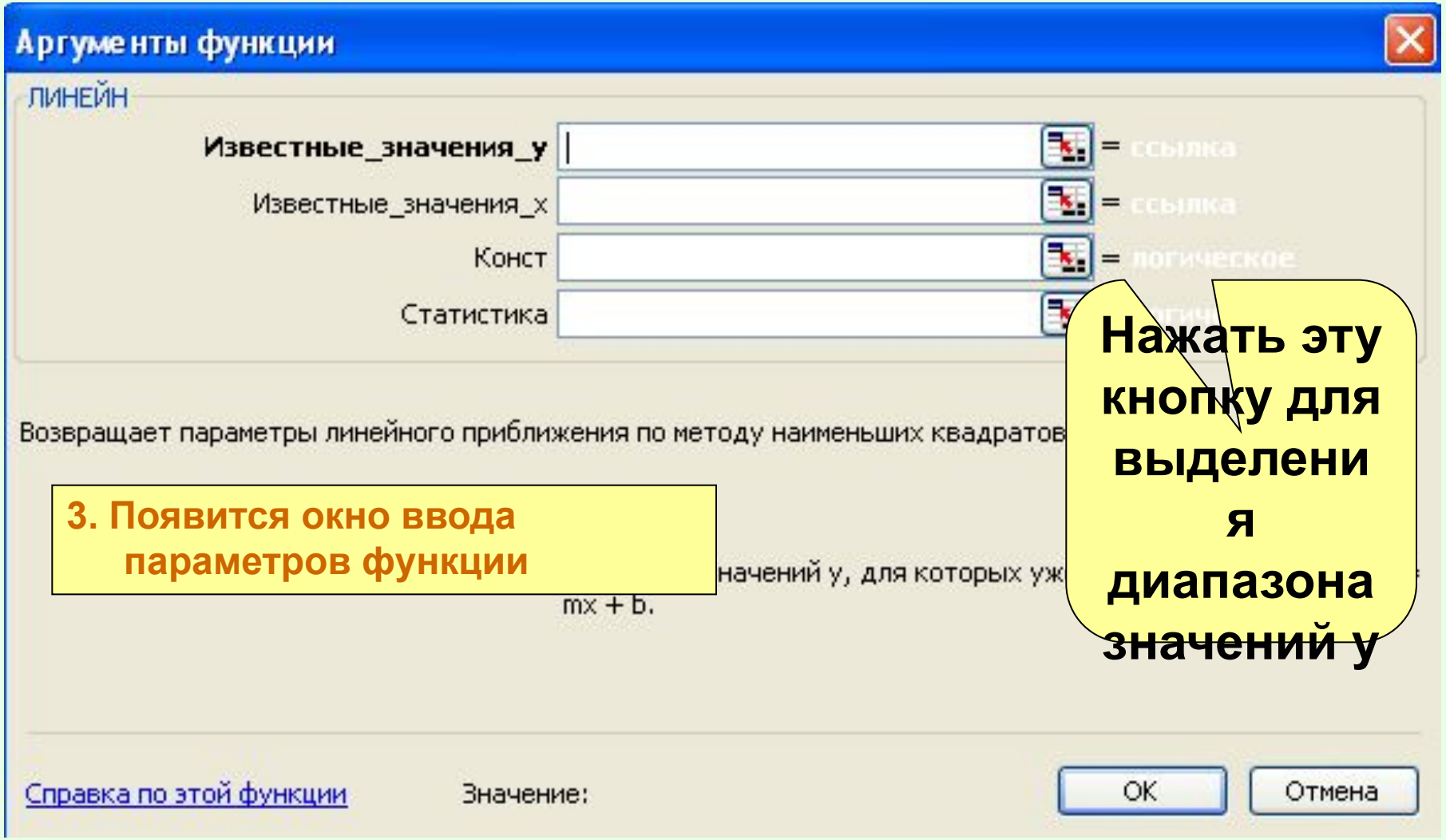
Значение коэффициента b	=		Значение коэффициента a
Среднеквадратическое отклонение b			Среднеквадратическое отклонение a
Коэффициент детерминации R^2			Среднеквадратическое отклонение y
F-статистика			Число степеней свободы
Регрессионаая сумма квадратов (RSS)			Остаточная сумма квадратов (ESS)

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



Линейный регрессионный анализ

Ход ввода параметров функции ЛИНЕЙН



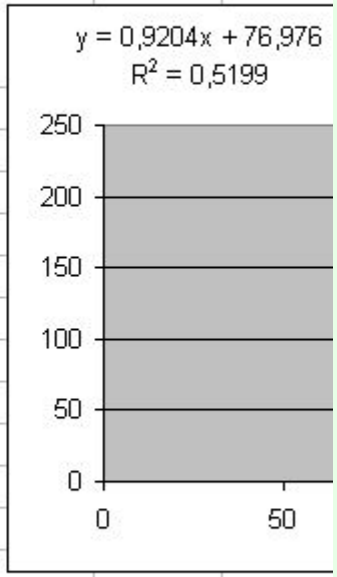
Линейный регрессионный анализ

Ход ввода параметров функции ЛИНЕЙН

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

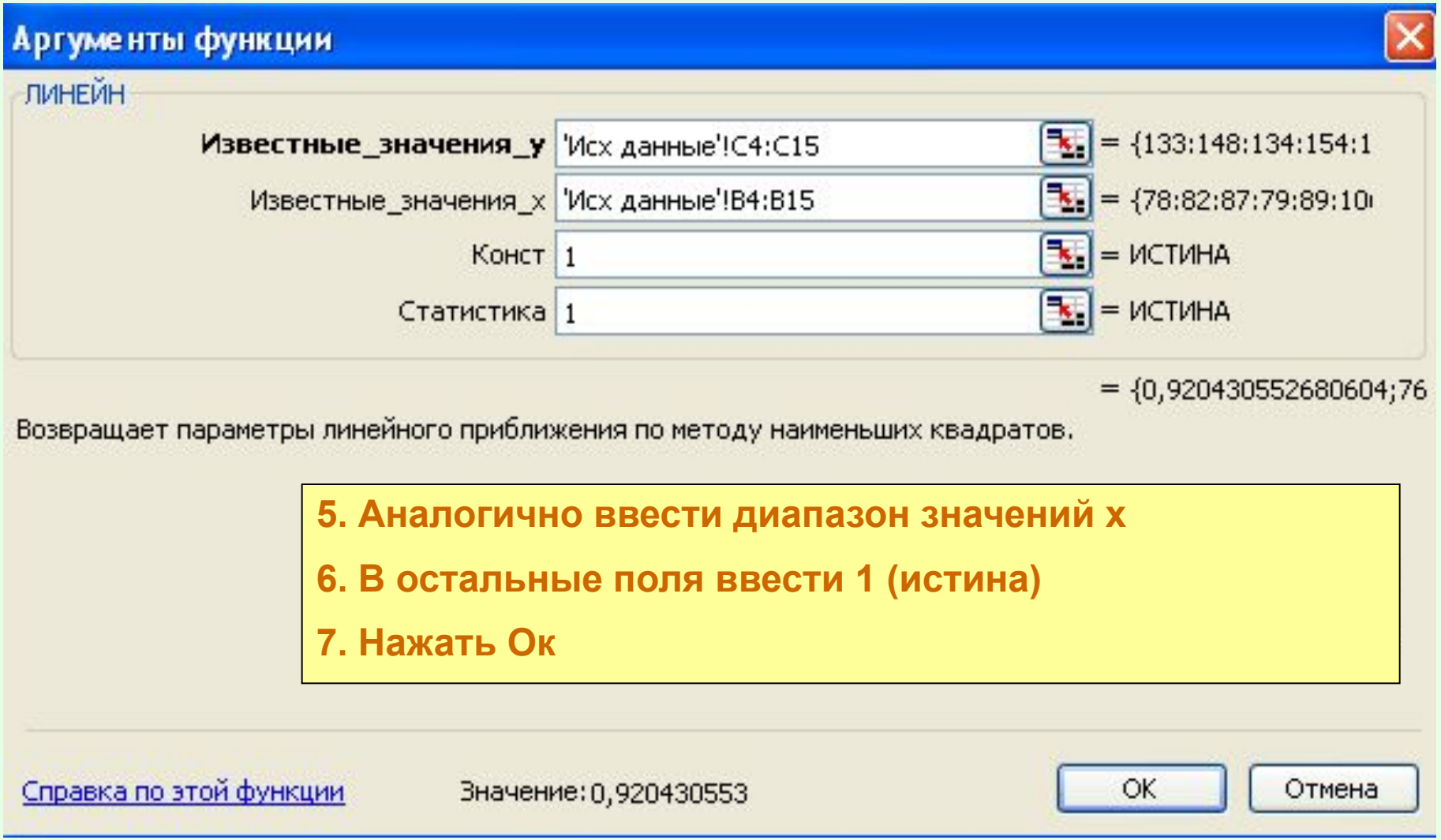
По территориям региона приводятся данные за 199X г.

Номер региона	Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., x	Среднедневная зарплата, руб., 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



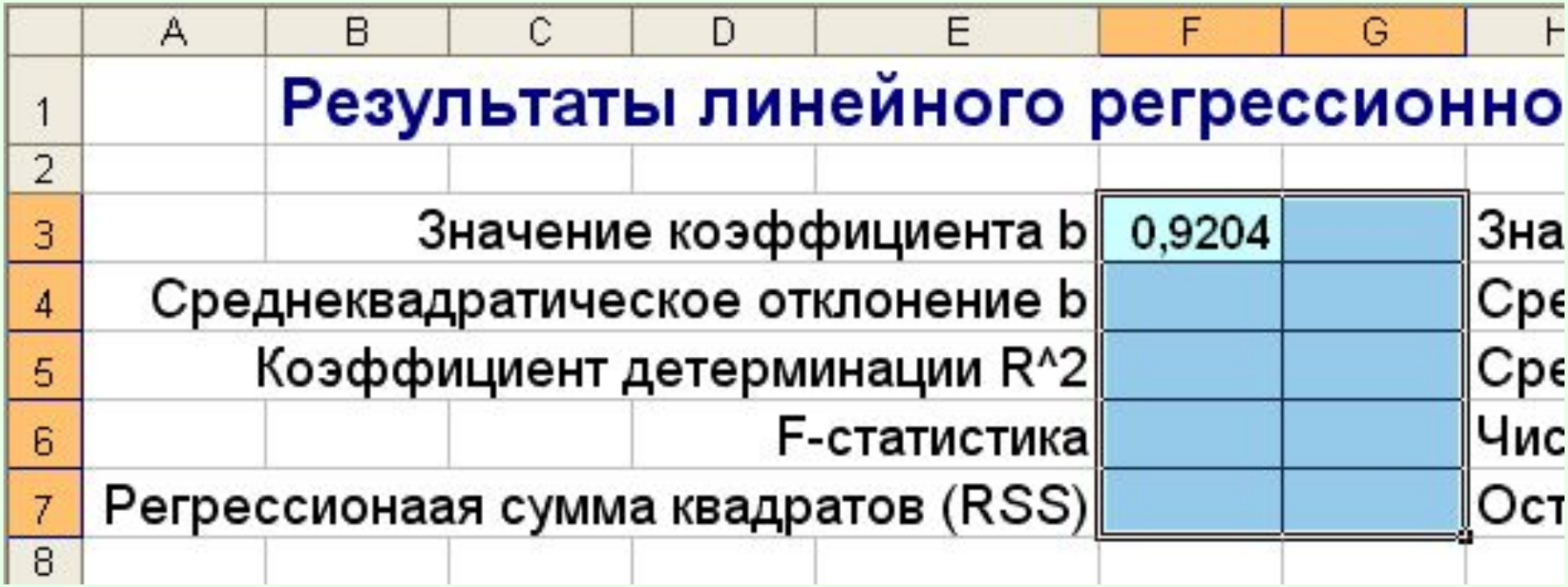
Линейный регрессионный анализ

Ход ввода параметров функции ЛИНЕЙН



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					Регрессионаая 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
Регрессионаая 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

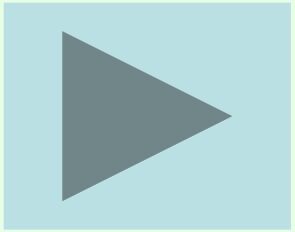
$$y = 0{,}9204x + 76{,}976$$

$$R^2 = 0{,}5199$$

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 5х2
3. Вызвать функцию ЛИНЕЙН и ввести ее параметры
4. Нажать F2, а затем CTRL+SHIFT+ENTER

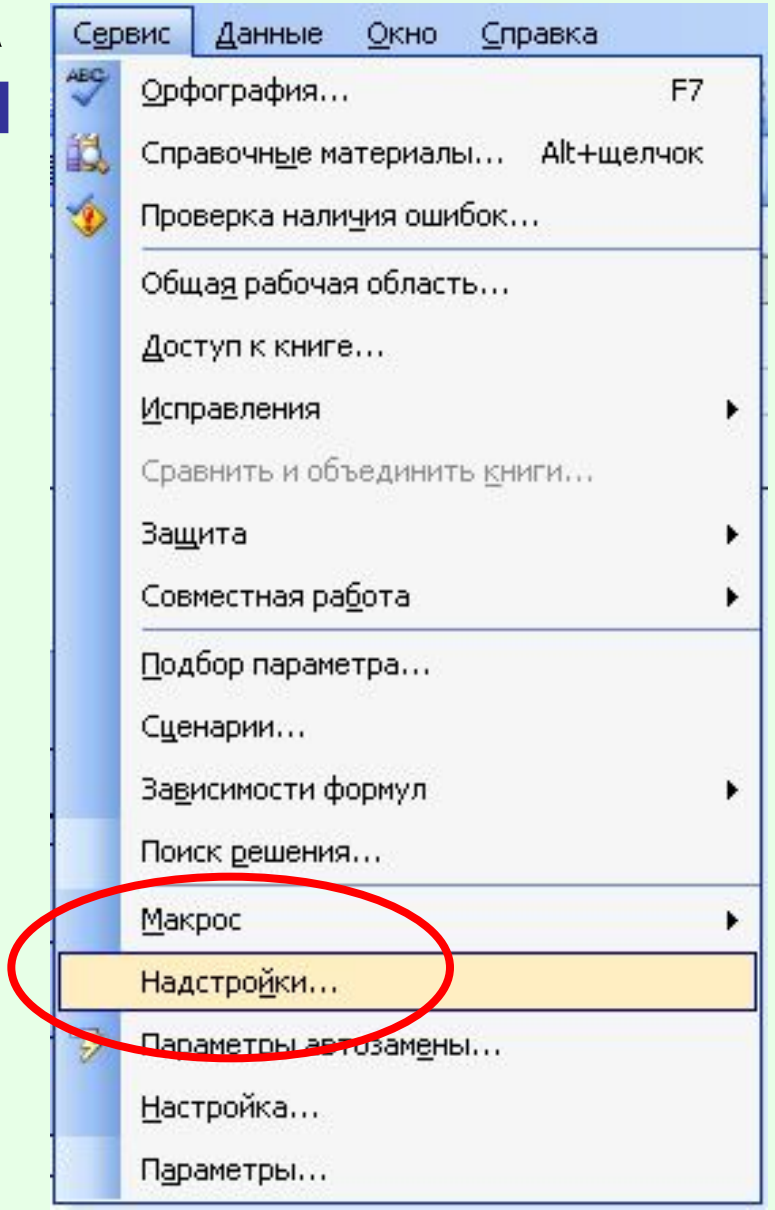
Записать в тетрадь

Линейный регрессионный анализ

Использование инструмента анализа данных РЕГРЕССИЯ

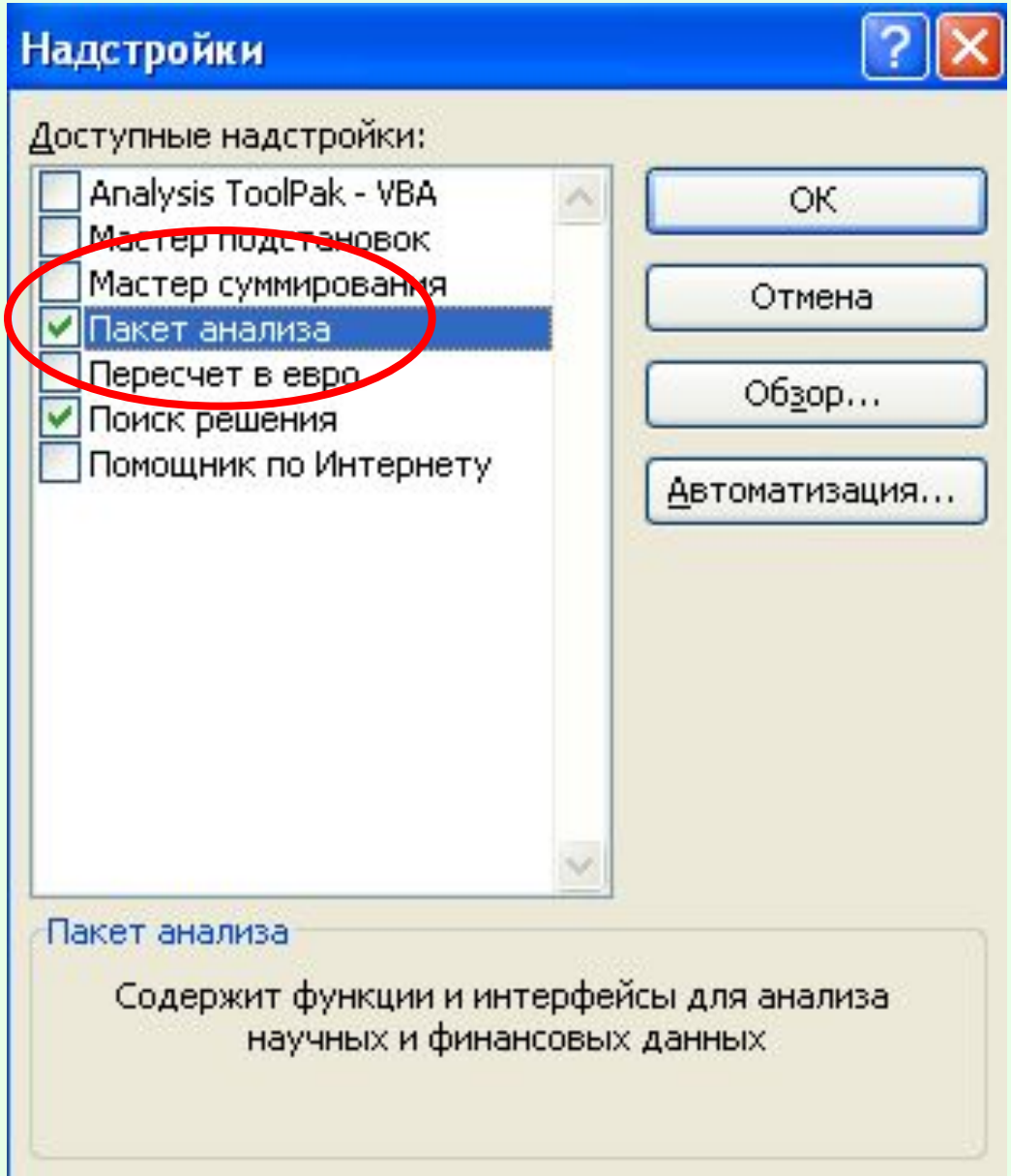
1. **Проверить доступ к пакету анализа**

   В главном меню последовательно выбрать СЕРВИС / НАДСТРОЙКИ.



Линейный регрессионный анализ

2. Установить флажок **ПАКЕТ АНАЛИЗА**
3. Нажать **Ок**



Линейный регрессионный анализ

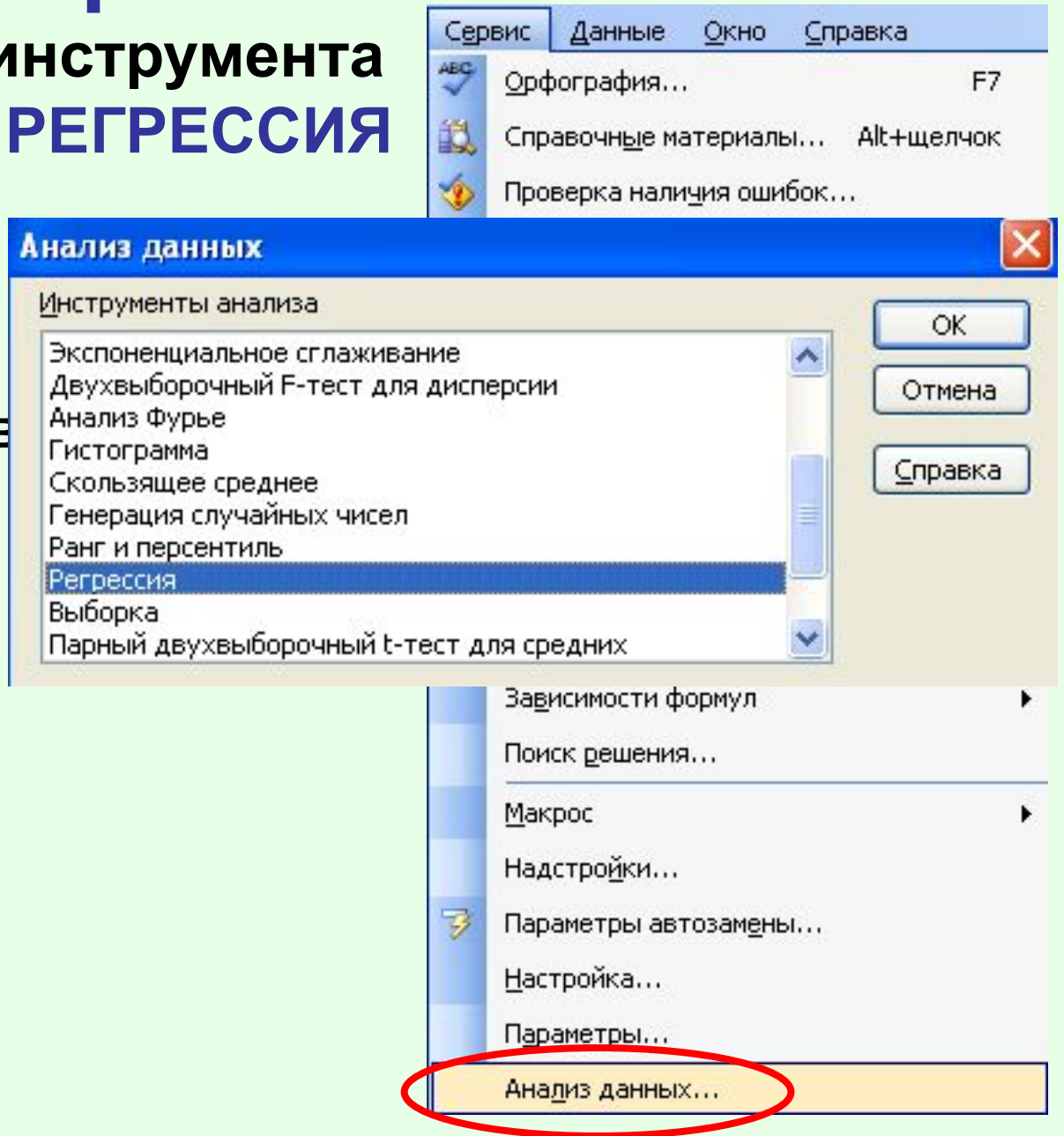
Использование инструмента анализа данных РЕГРЕССИЯ

4. В главном меню в

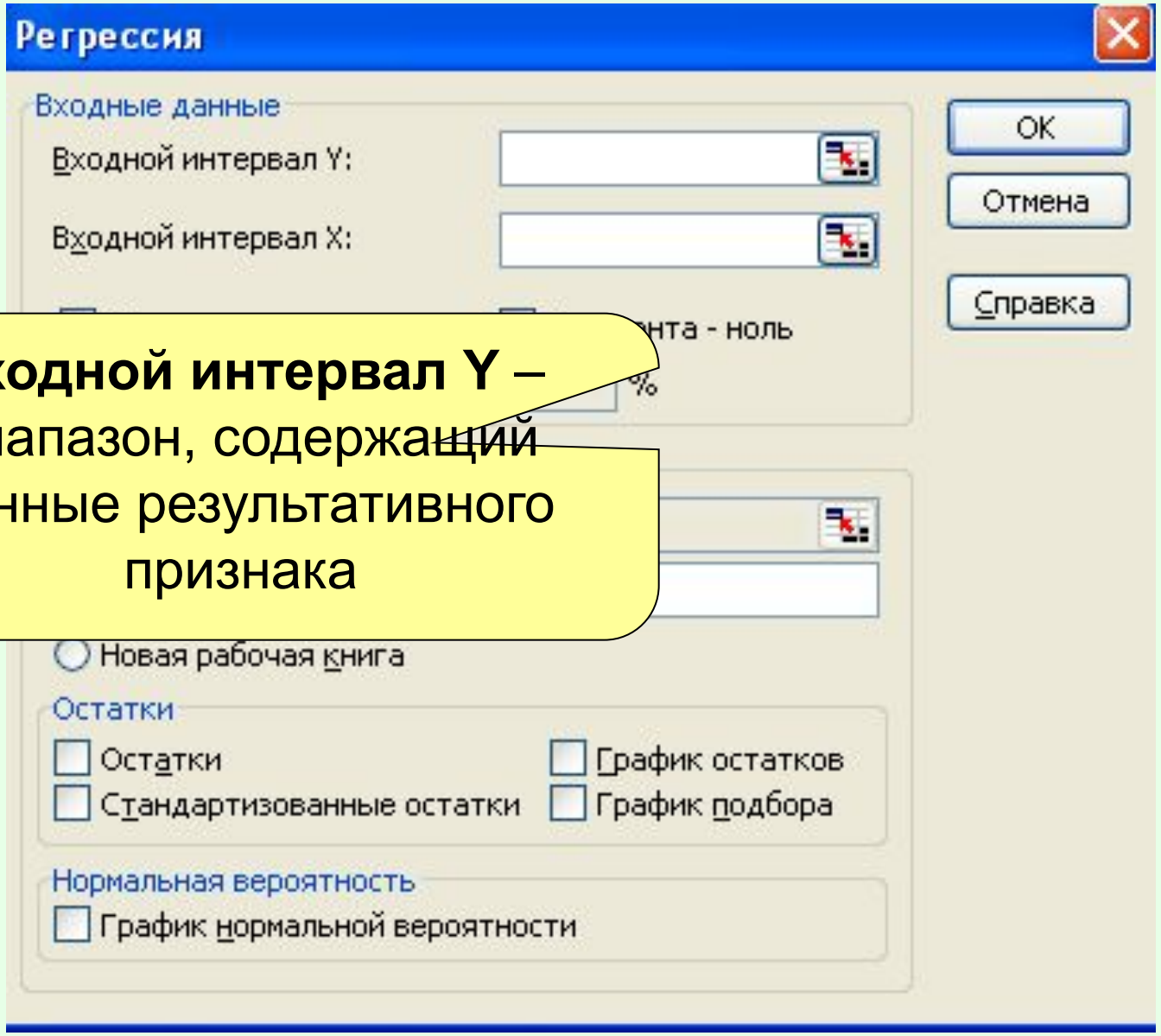
СЕРВИС/

АНАЛИЗ ДАННЫХ/

РЕГРЕССИЯ

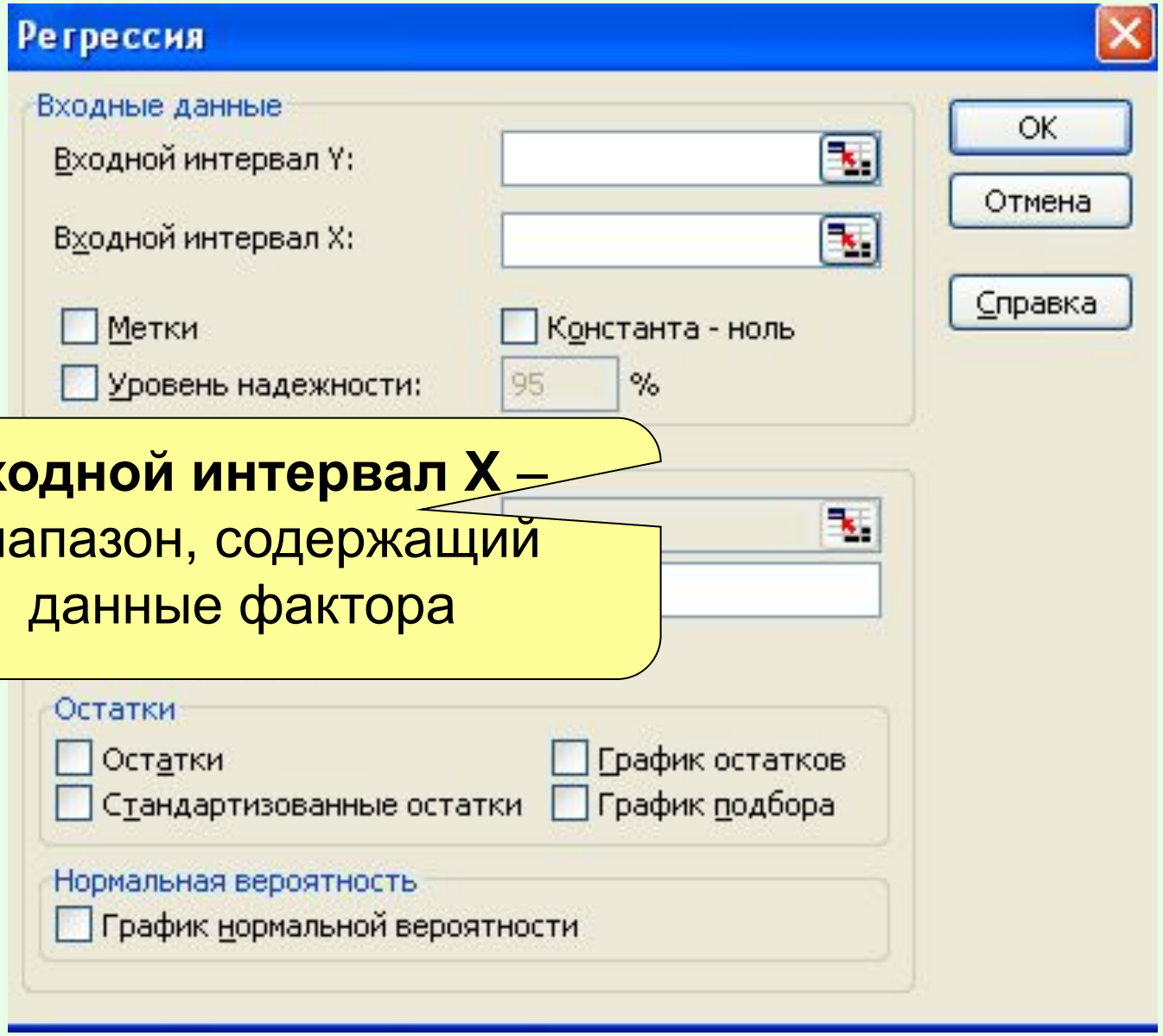


Использование инструмента РЕГРЕССИЯ



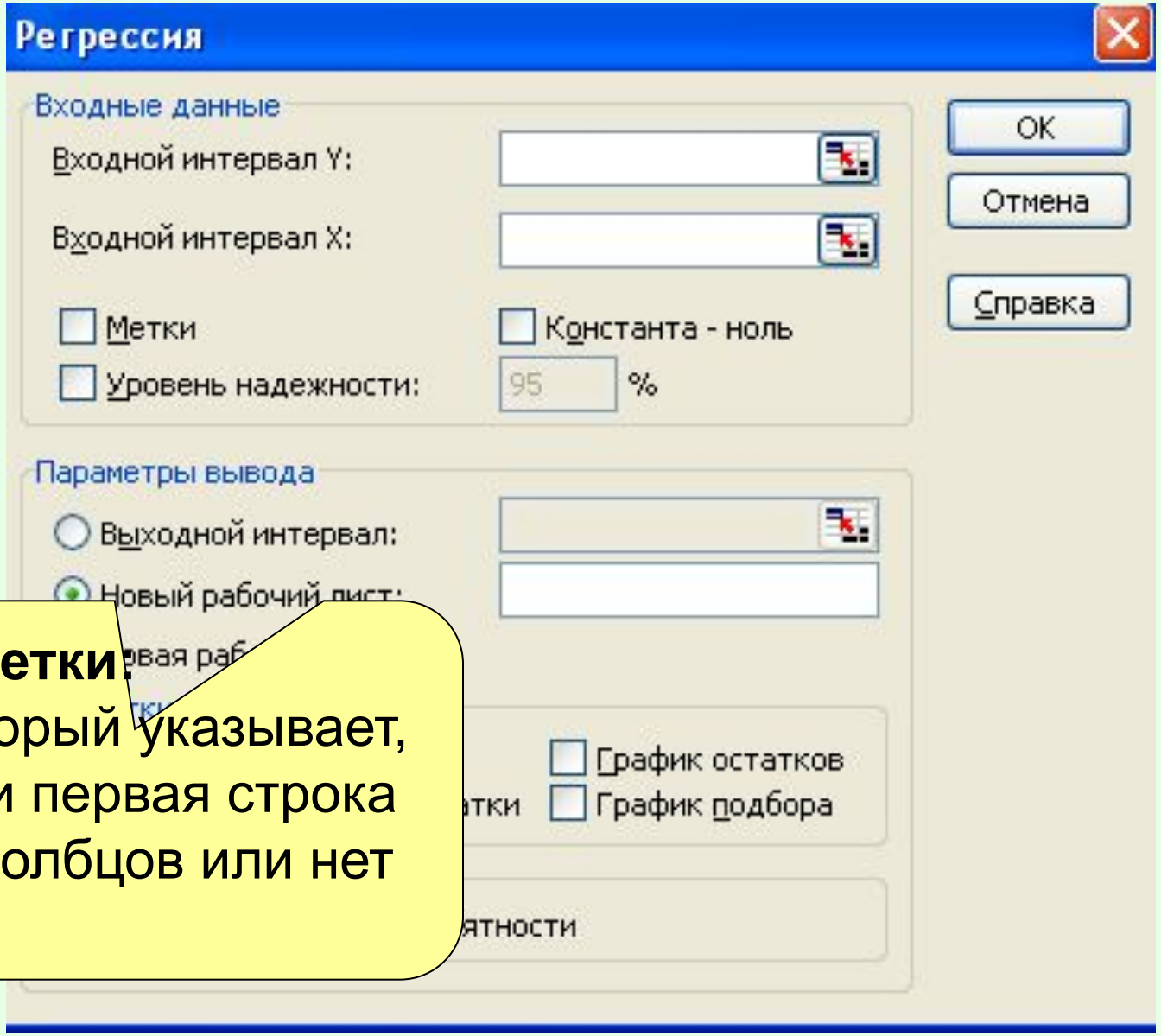
Входной интервал Y – диапазон, содержащий данные результативного признака

Использование инструмента РЕГРЕССИЯ



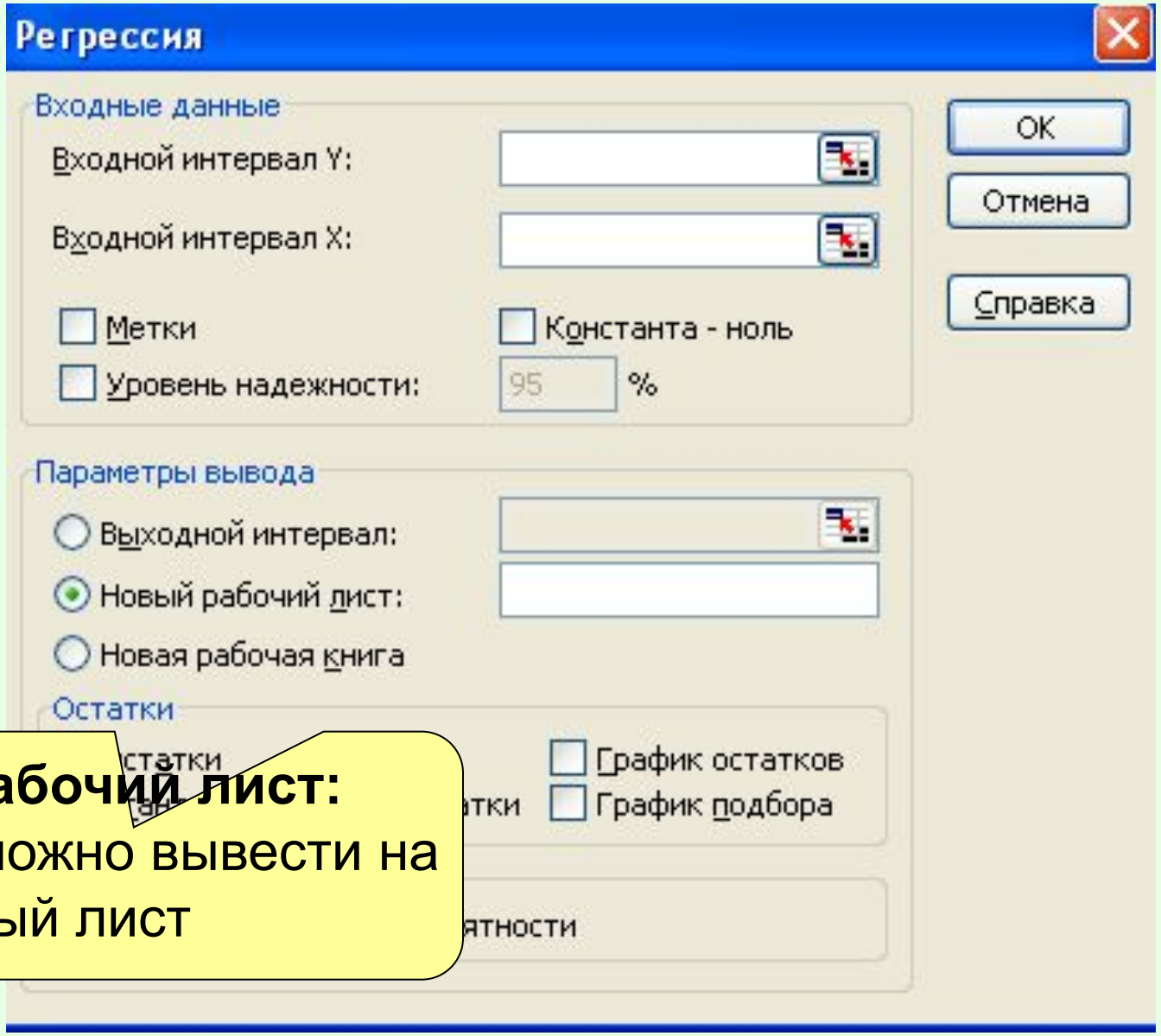
Входной интервал X – диапазон, содержащий данные фактора

Использование инструмента РЕГРЕССИЯ



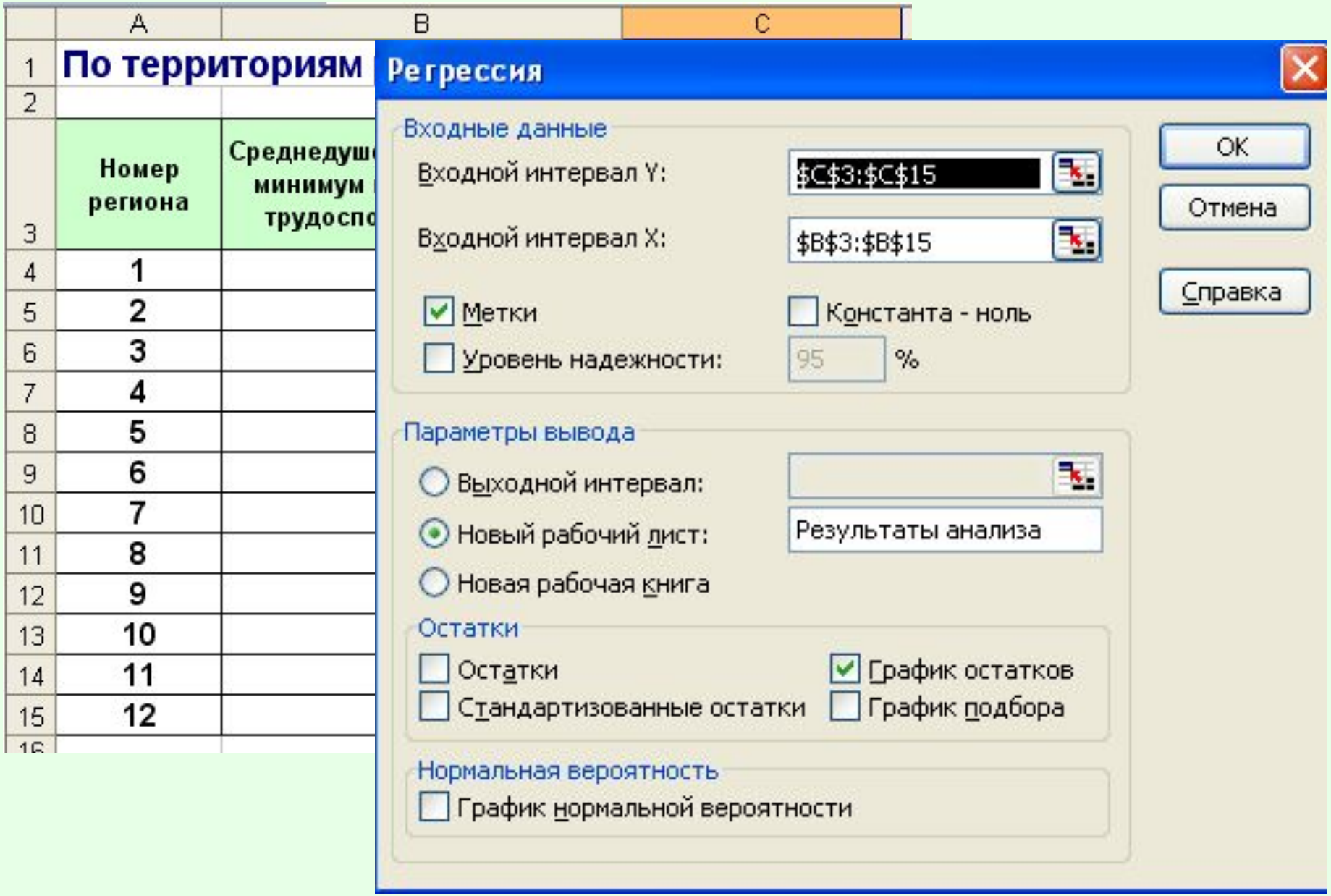
Метки:
флажок, который указывает, содержит ли первая строка названия столбцов или нет

Использование инструмента РЕГРЕССИЯ



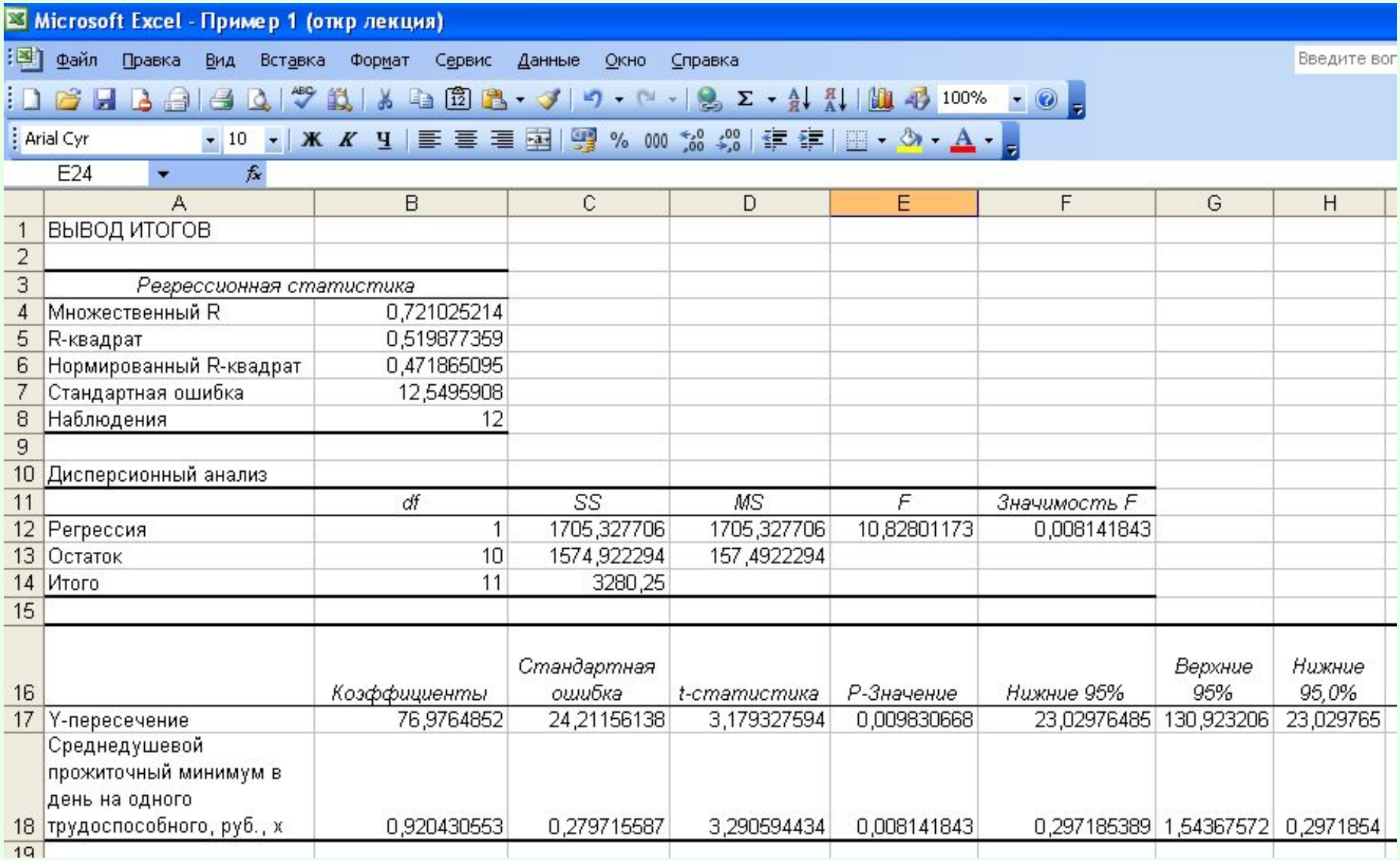
Новый рабочий лист: результаты можно вывести на новый лист

Использование инструмента РЕГРЕССИЯ



	A	B	C
1	По территориям		
2			
3	Номер региона	Среднедуш… минимум … трудоспо…	
4	1		
5	2		
6	3		
7	4		
8	5		
9	6		
10	7		
11	8		
12	9		
13	10		
14	11		
15	12		

Использование инструмента РЕГРЕССИЯ



	A	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		*df*	*SS*	*MS*	*F*	*Значимость F*		
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		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*
17	Y-пересечение	76,9764852	24,21156138	3,179327594	0,009830668	23,02976485	130,923206	23,029765
18	Среднедушевой прожиточный минимум в день на одного трудоспособного, руб., x	0,920430553	0,279715587	3,290594434	0,008141843	0,297185389	1,54367572	0,2971854
19								

Использование инструмента РЕГРЕССИЯ

Регрессионная статистика	
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

Регрессионная статистика	
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				
	df	*SS*	*MS*	*F*
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

	Коэффициенты	*Стандартная ошибка*	*t-статистика*
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

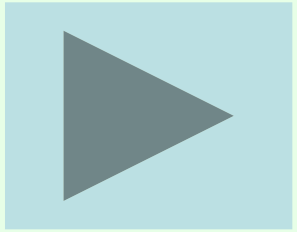
Параметр **a**: **76.976**

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

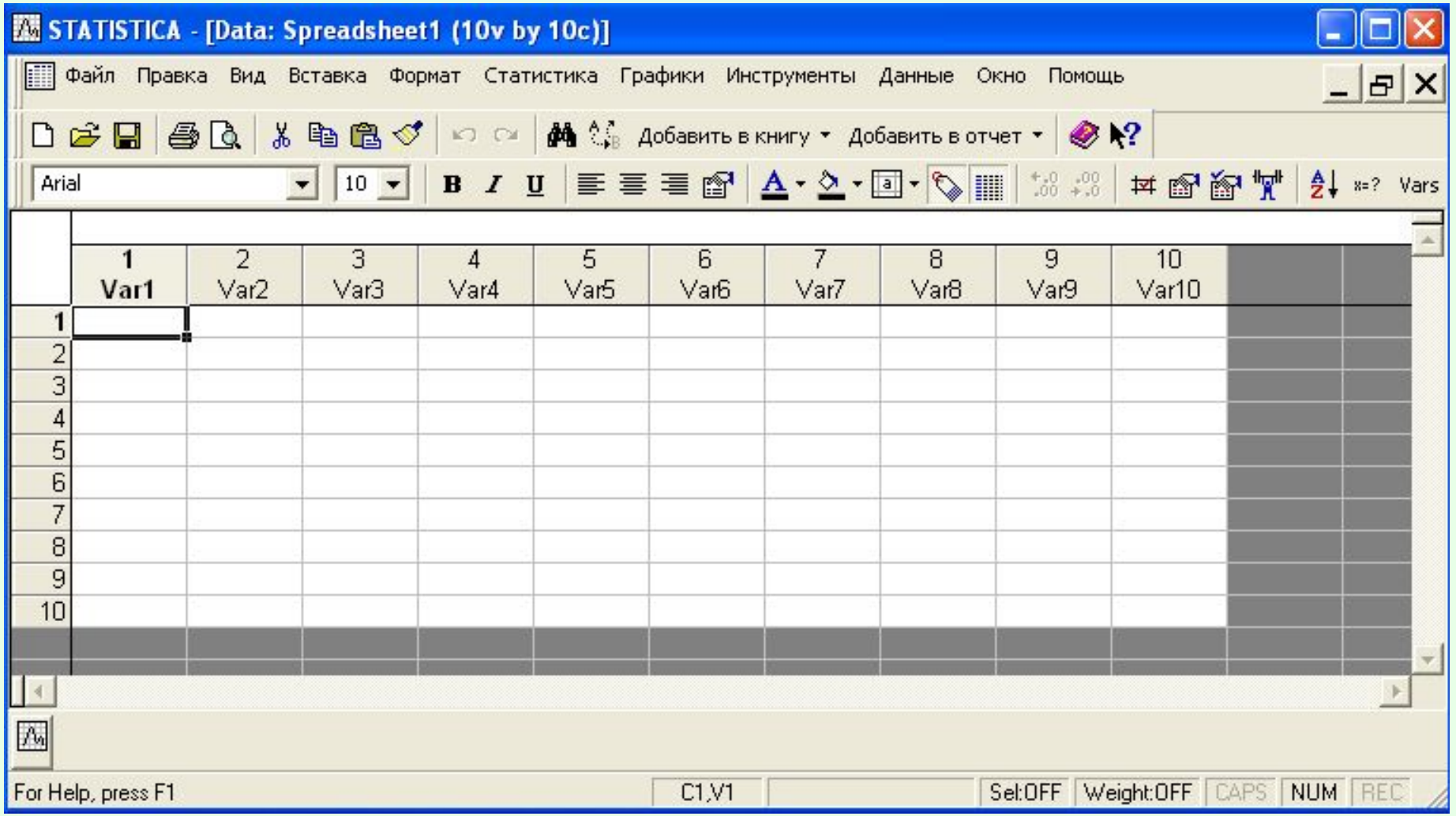
3) Использование инструмента РЕГРЕССИЯ

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

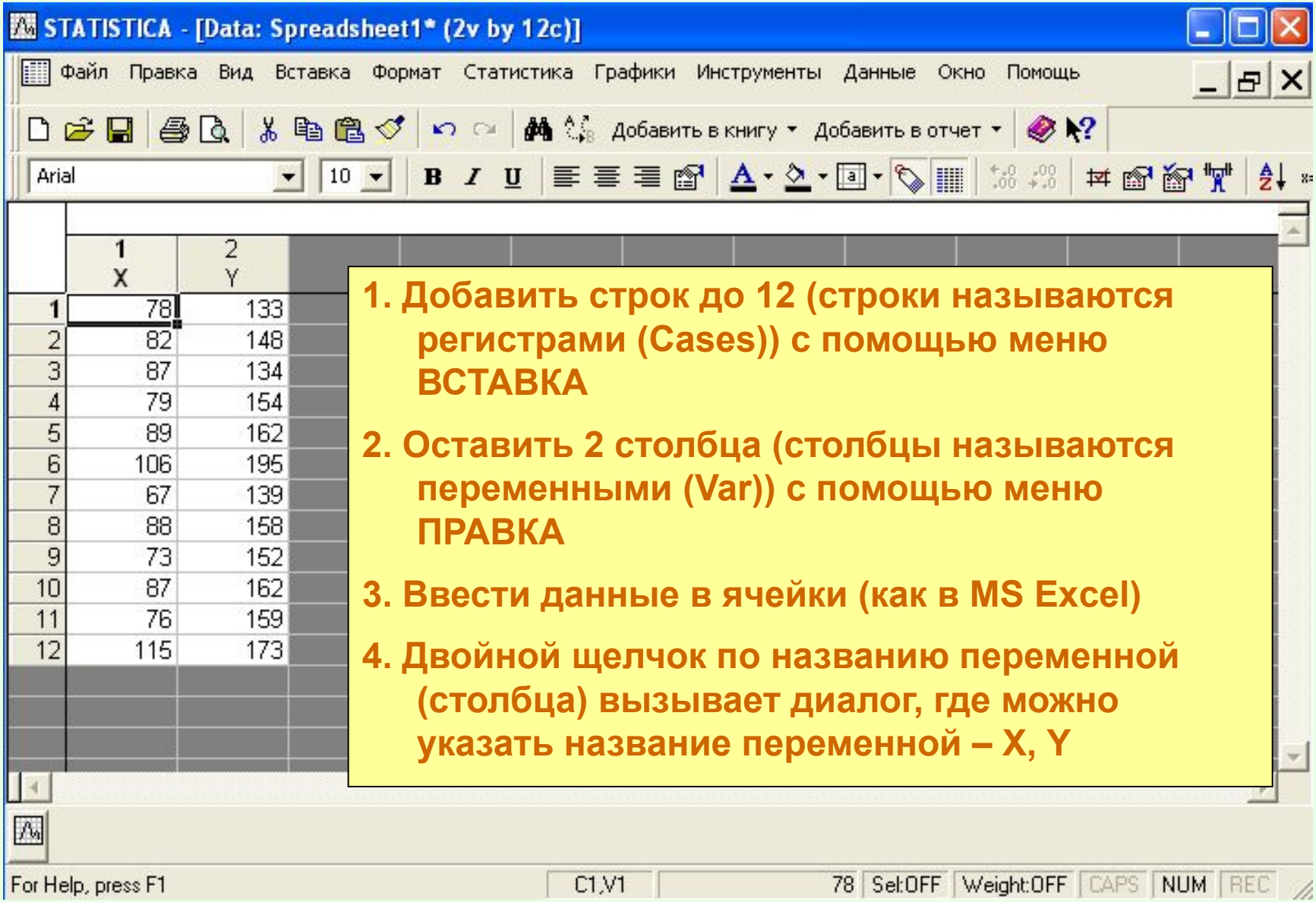
Записать в тетрадь

2. Использование STATISTICA для решения задач по теме «Линейная парная регрессия»

Окно программы STATISTICA

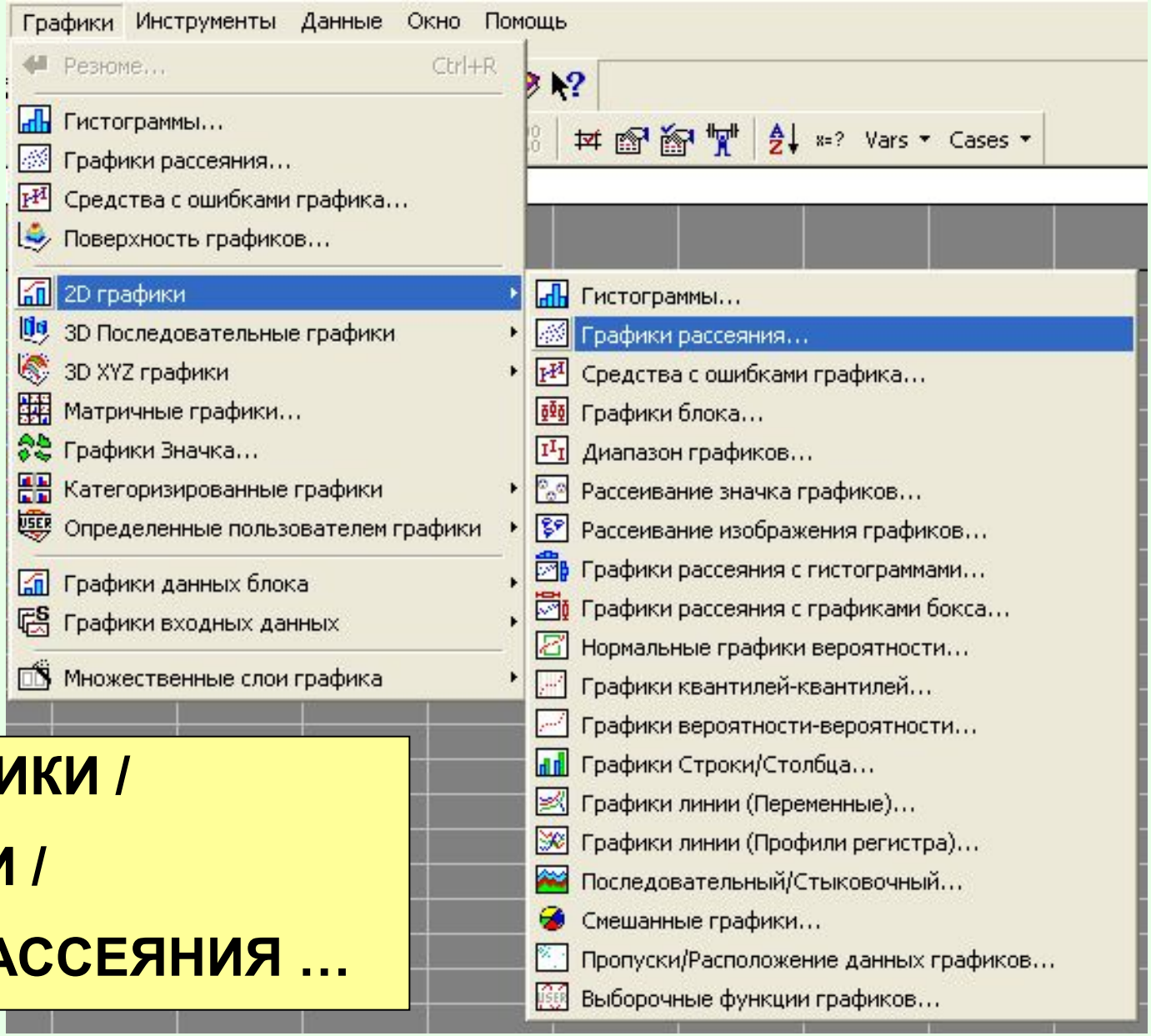


Ввод данных для анализа



	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

Построение графика рассеяния переменных

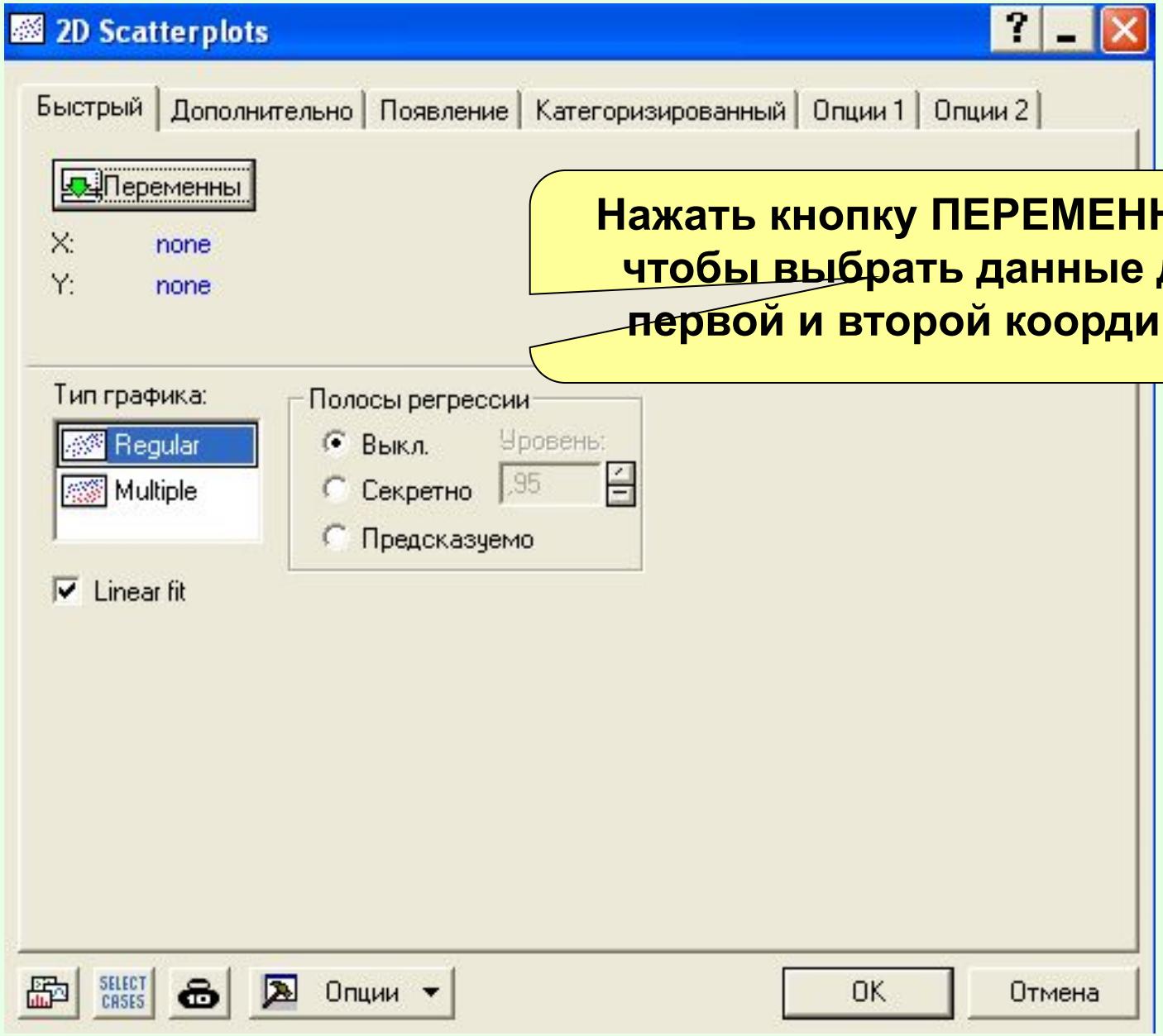


Меню ГРАФИКИ /

2D ГРАФИКИ /

ГРАФИКИ РАССЕЯНИЯ …

Построение графика рассеяния переменных

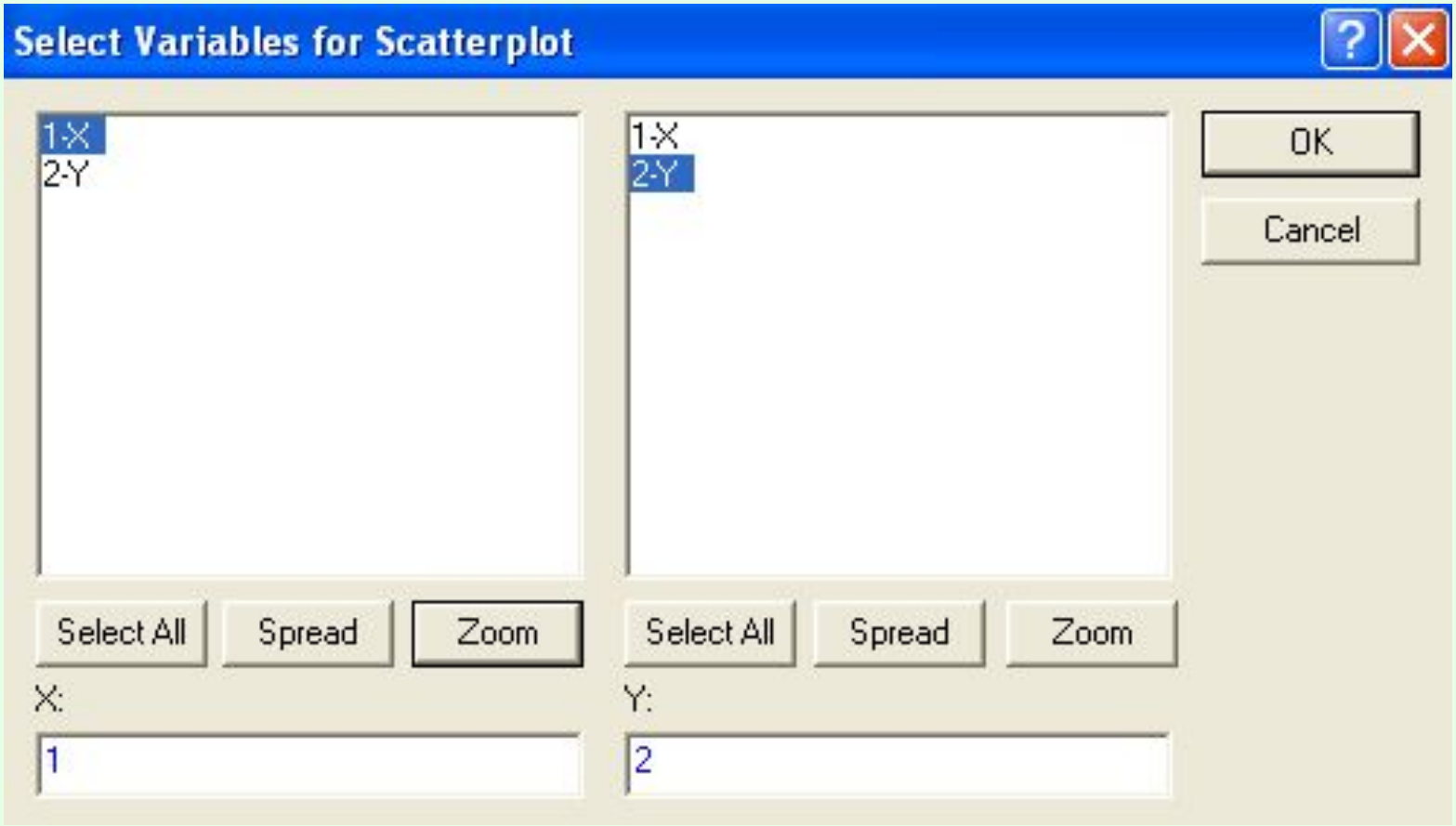


Нажать кнопку ПЕРЕМЕННЫЕ, чтобы выбрать данные для первой и второй координат

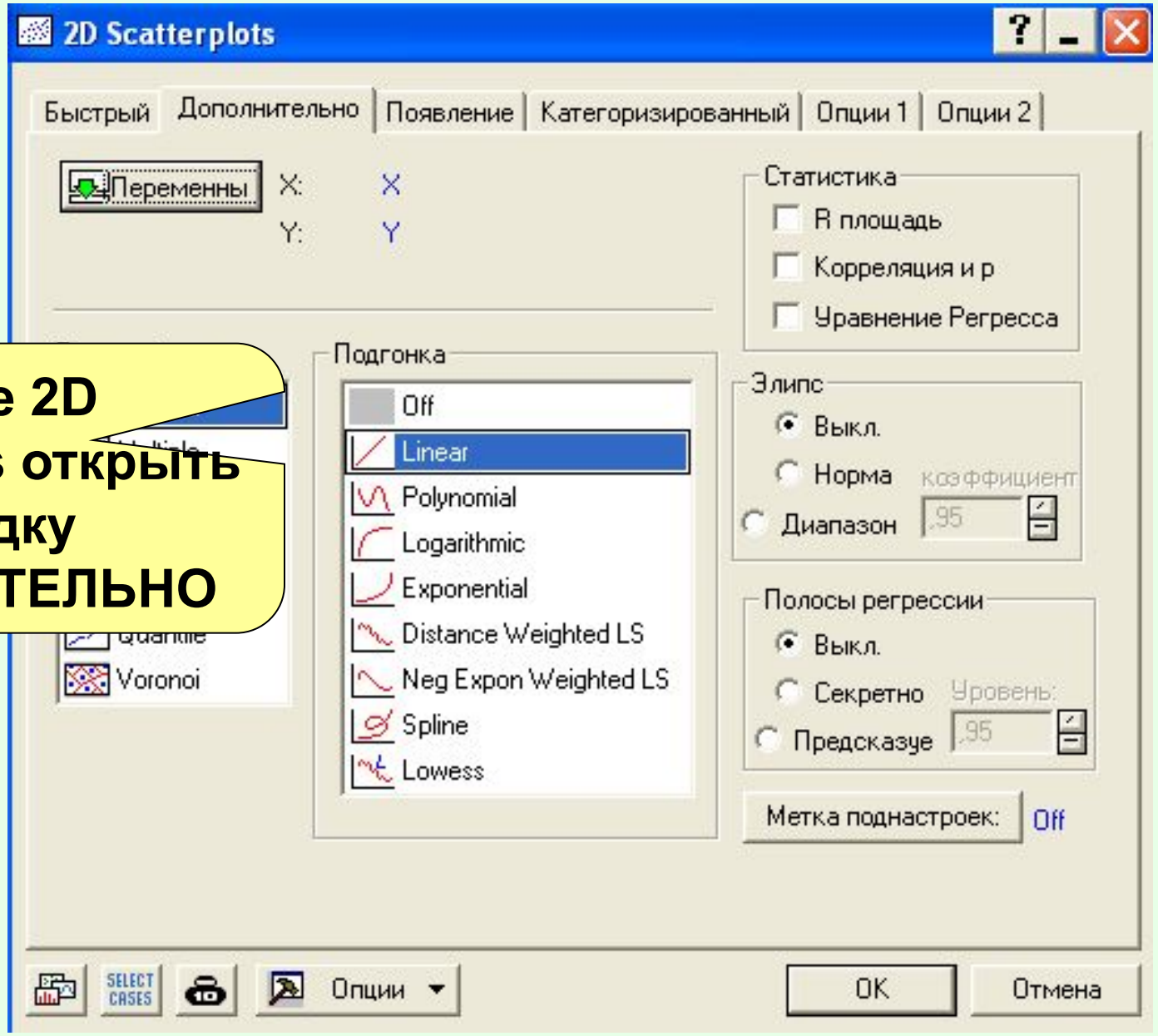
Построение графика рассеяния переменных

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

- Нажать Ок



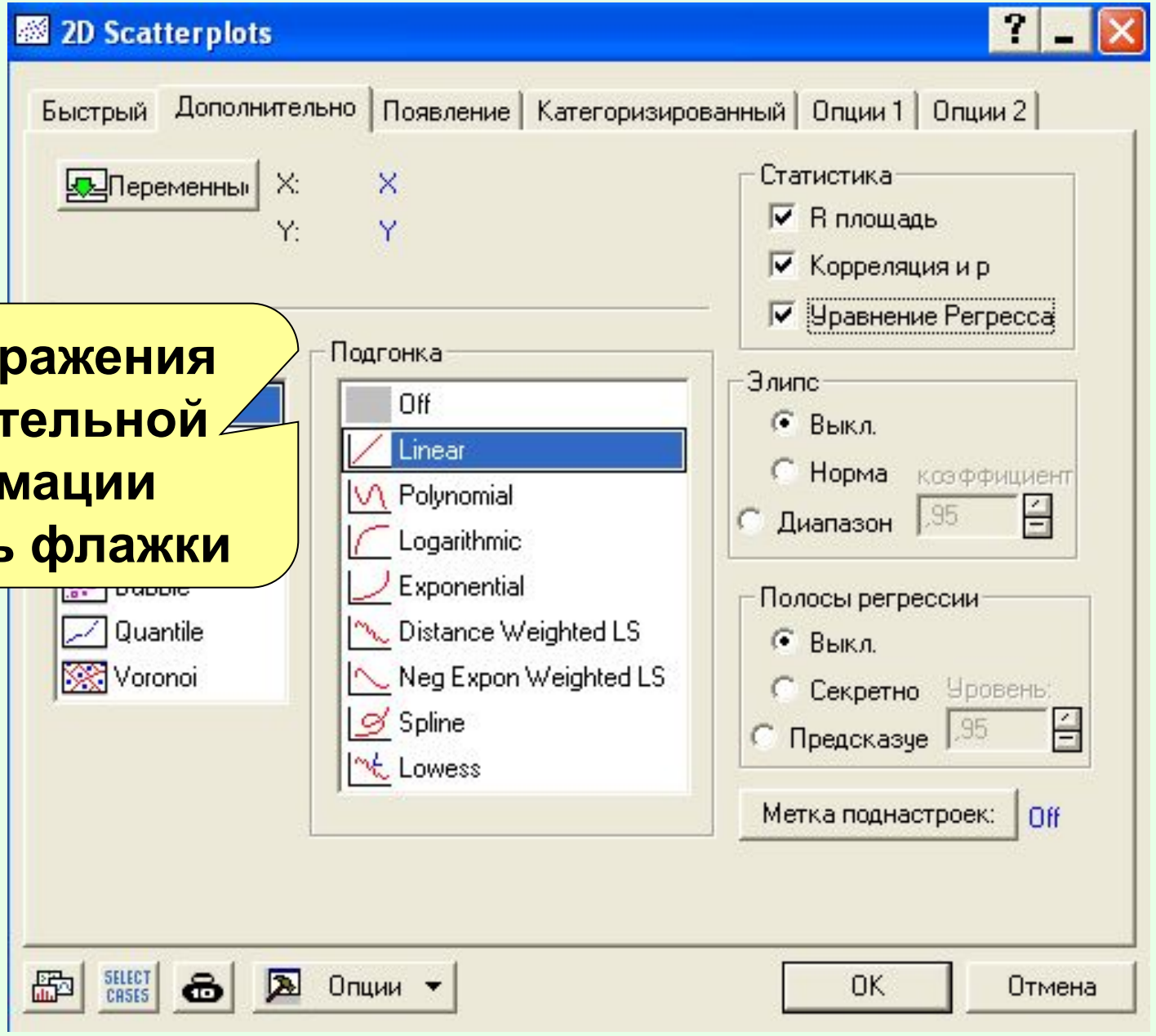
Построение графика рассеяния переменных



В окне 2D Scatterplots открыть вкладку ДОПОЛНИТЕЛЬНО

Построение графика рассеяния переменных

Для отображения дополнительной информации поставить флажки



Построение графика рассеяния переменных

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

Линейная регрессия в STATISTICA

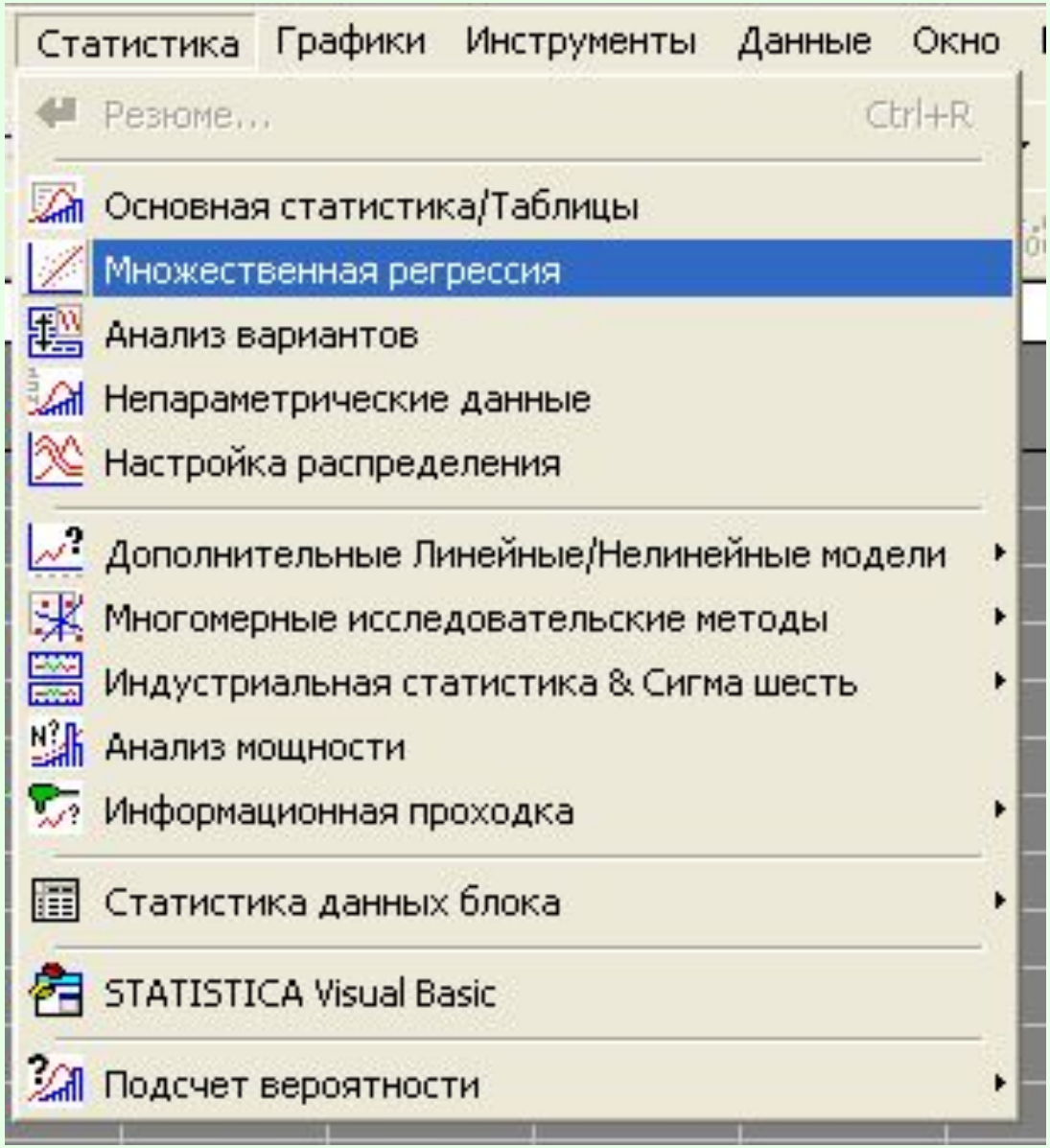
1) Построение диаграммы рассеяния

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

Записать в тетрадь

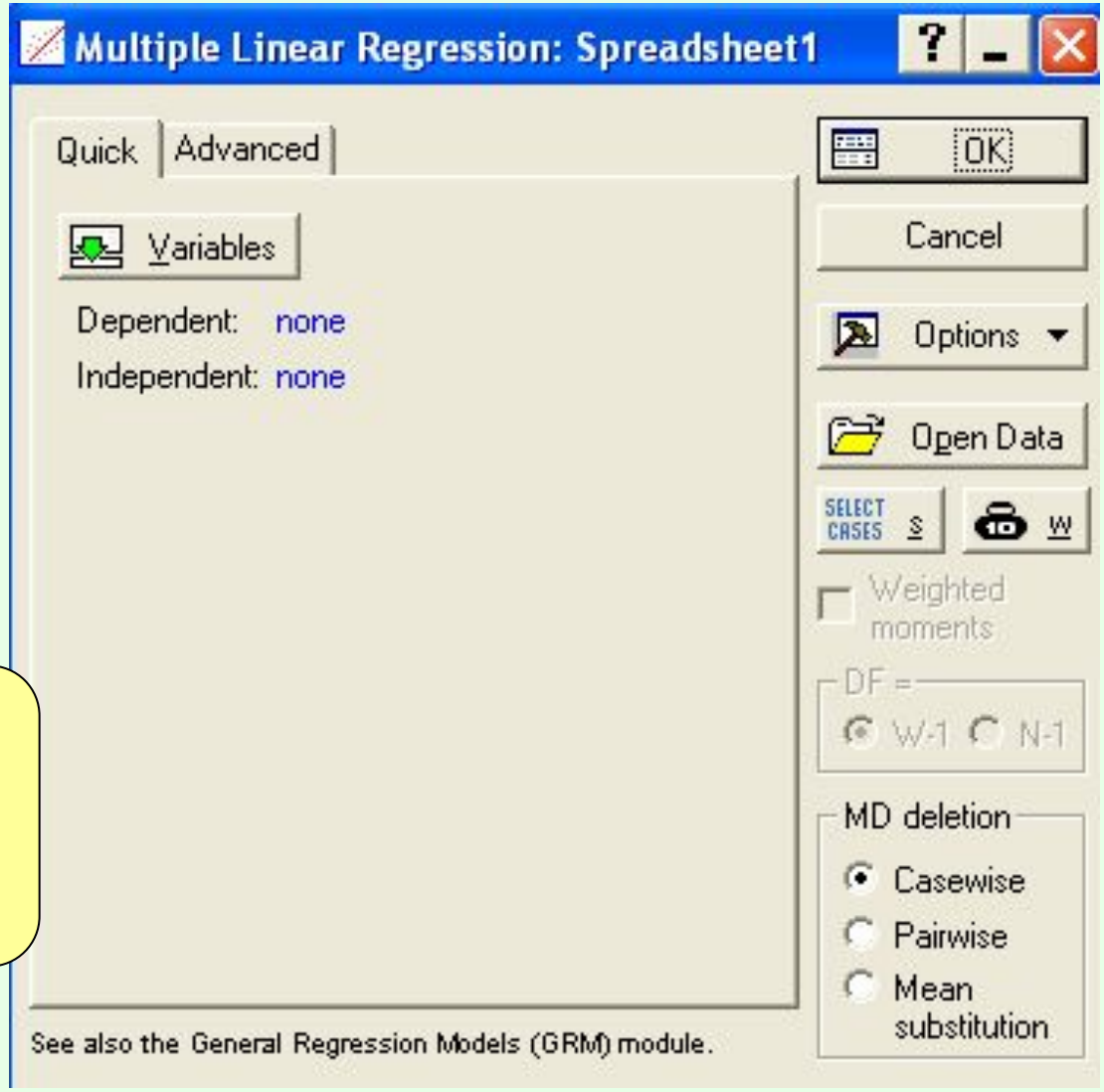
Регрессионный анализ

- Меню СТАТИСТИКА
- Множественная регрессия



Регрессионный анализ

Нажать кнопку Variables (Переменные)

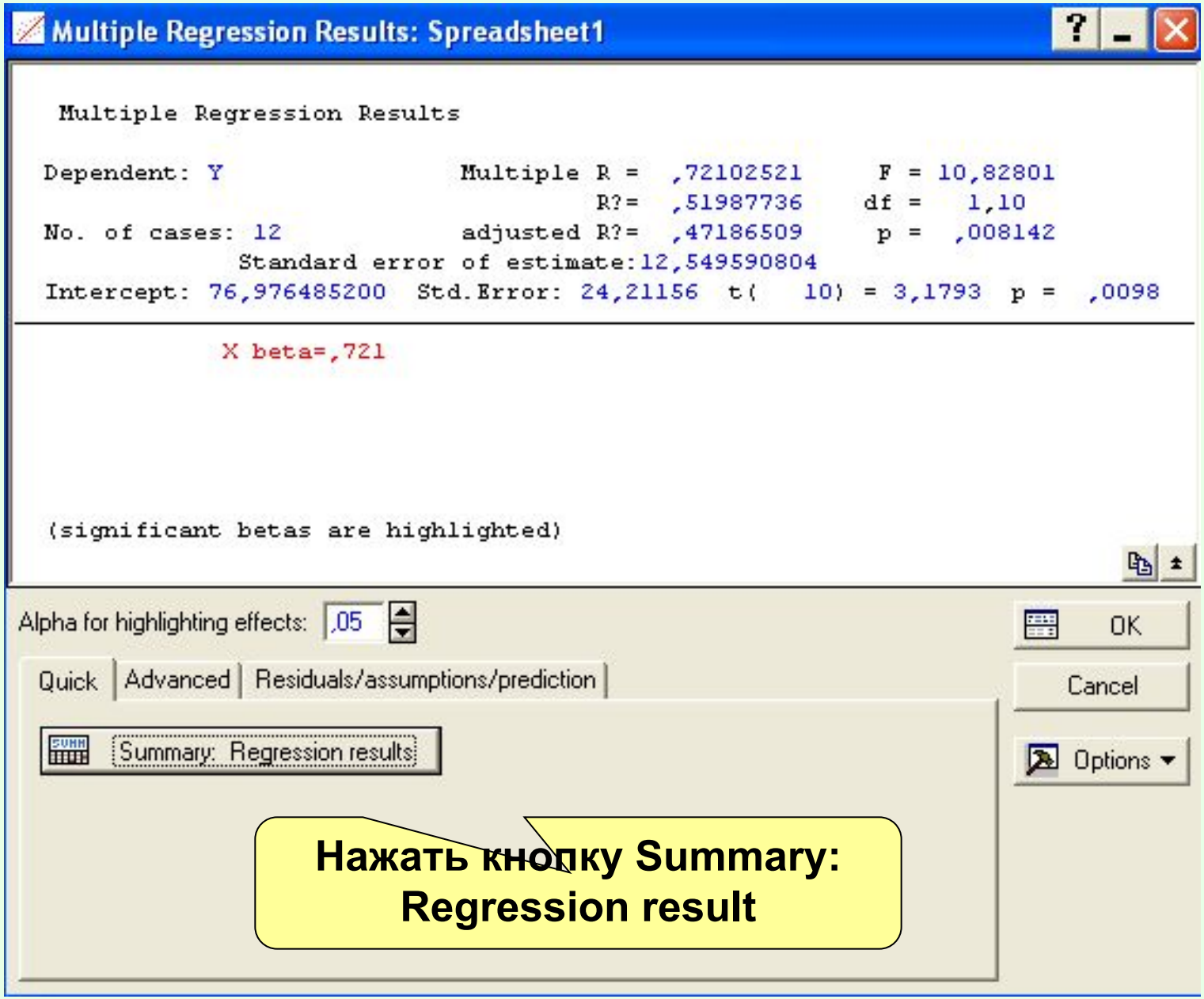


Регрессионный анализ

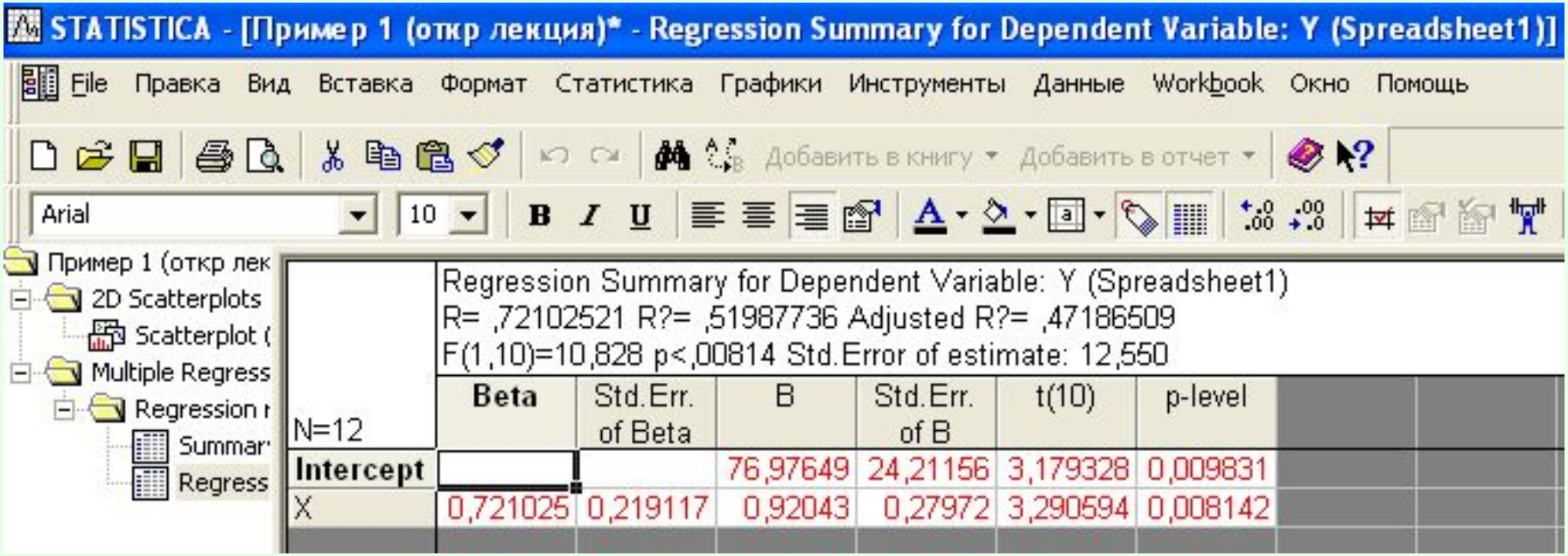
Слева указать переменную Y (Dependent – зависимая переменная)

Справа указать переменную X (Independent – независимая переменная)

Регрессионный анализ



Регрессионный анализ



Regression Summary for Dependent Variable: Y (Spreadsheet1)
R= ,72102521 R?= ,51987736 Adjusted R?= ,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

Записать в тетрадь