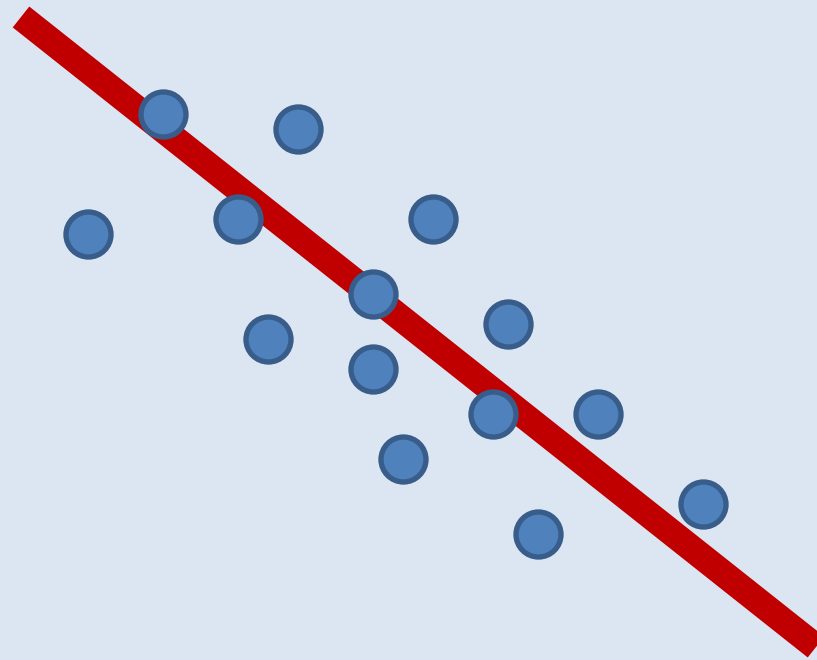


# регрессионный



# анализ

# регрессионный анализ - введение

**коэффициент корреляции** –  
степень связи в вариации  
двух переменных величин  
(мера тесноты этой связи)

**метод регрессии** –  
позволяет судить как количественно  
меняется одна величина по мере  
изменения другой

# регрессионный анализ - введение

метод регрессии со временем  
становился все более востребованным,  
чем корреляционный анализ,  
т.к. является более информативным

сегодня исследователи все чаще  
применяют оба метода в комплексе

# регрессионный анализ - введение

задача регрессионного анализа:

установить, как **количественно** меняется одна величина при изменении другой на единицу

регрессия может быть **двухсторонней**, т.к. переменных две –  $x$  и  $y$ .

# регрессионный анализ - введение

результатом регрессионного анализа  
считается:

- ✓ расчет коэффициента регрессии
- ✓ построения теоретической  
линии регрессии
- ✓ расчета уравнения регрессии
- ✓ построение эмпирической  
линии регрессии

# линейная регрессия

Общая формула уравнения  
функциональной зависимости

$$y=f(x),$$

где  $y$  рассматривают в качестве  
зависимой переменной, или функции от  
другой – независимой – переменной  
величины  $x$ , называемой аргументом

# линейная регрессия

**Регрессия** – это изменение среднего уровня одного признака при изменении другого

- не зависимость отдельных значений  **$y$**  от величины  **$x$**

+ зависимость частных средних  $\bar{y}_x$  от значений  $x$ .

# линейная регрессия

показатели регрессии выражают  
корреляционную связь **двусторонне**:

изменение усредненных значений  $\bar{y}_x$   
признака  $Y$  при изменении значений  $x_i$

изменение средних значений  $\bar{x}_y$   
признака  $X$  по измененным значениям  $y_i$



# линейная регрессия

важнейшим **результатом**  
**регрессионного анализа** служат:

- ✓ оценки значимости параметров регрессионного уравнения
- ✓ показатель силы влияния изучаемого фактора на признак

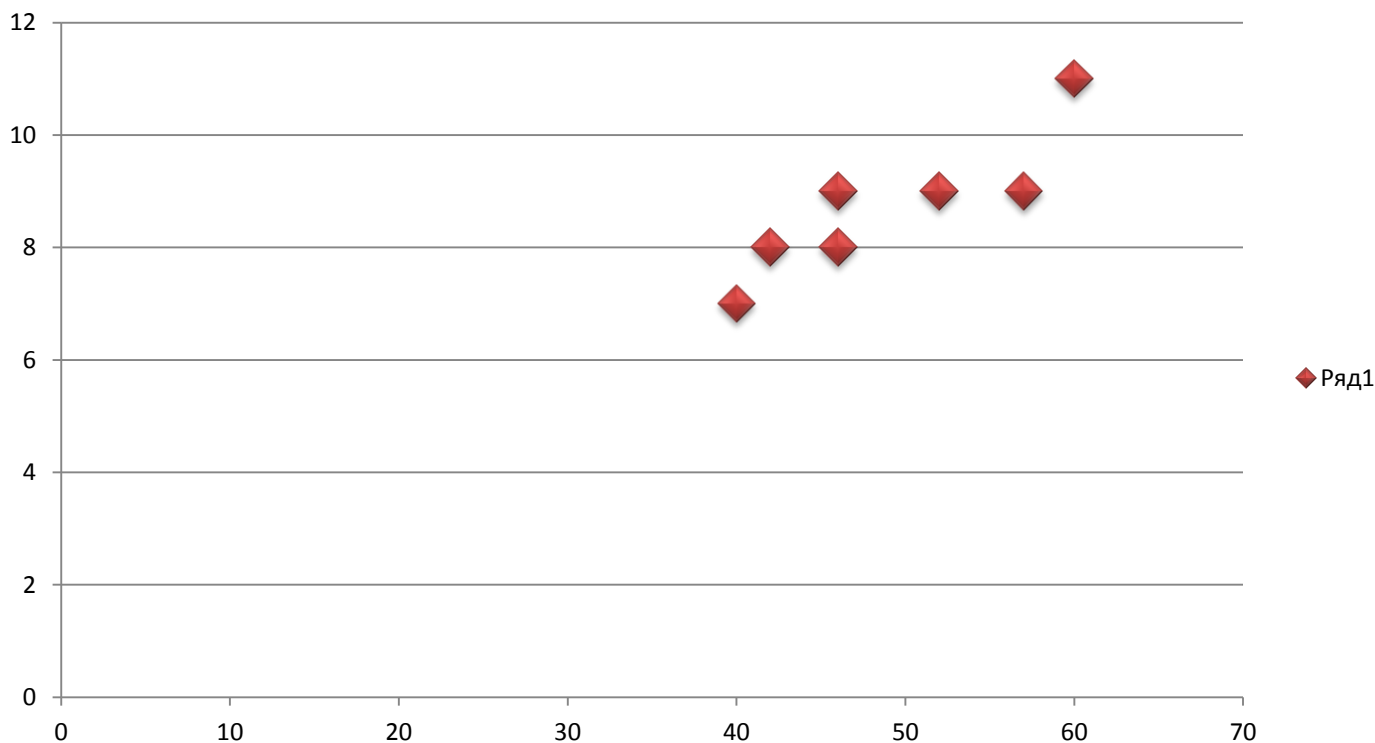
# линейная регрессия

главный результат  
использования линейной  
регрессии?

с помощью полученной в расчетах  
уравнении можно описывать и  
прогнозировать явления и процессы  
в проводимом исследовании

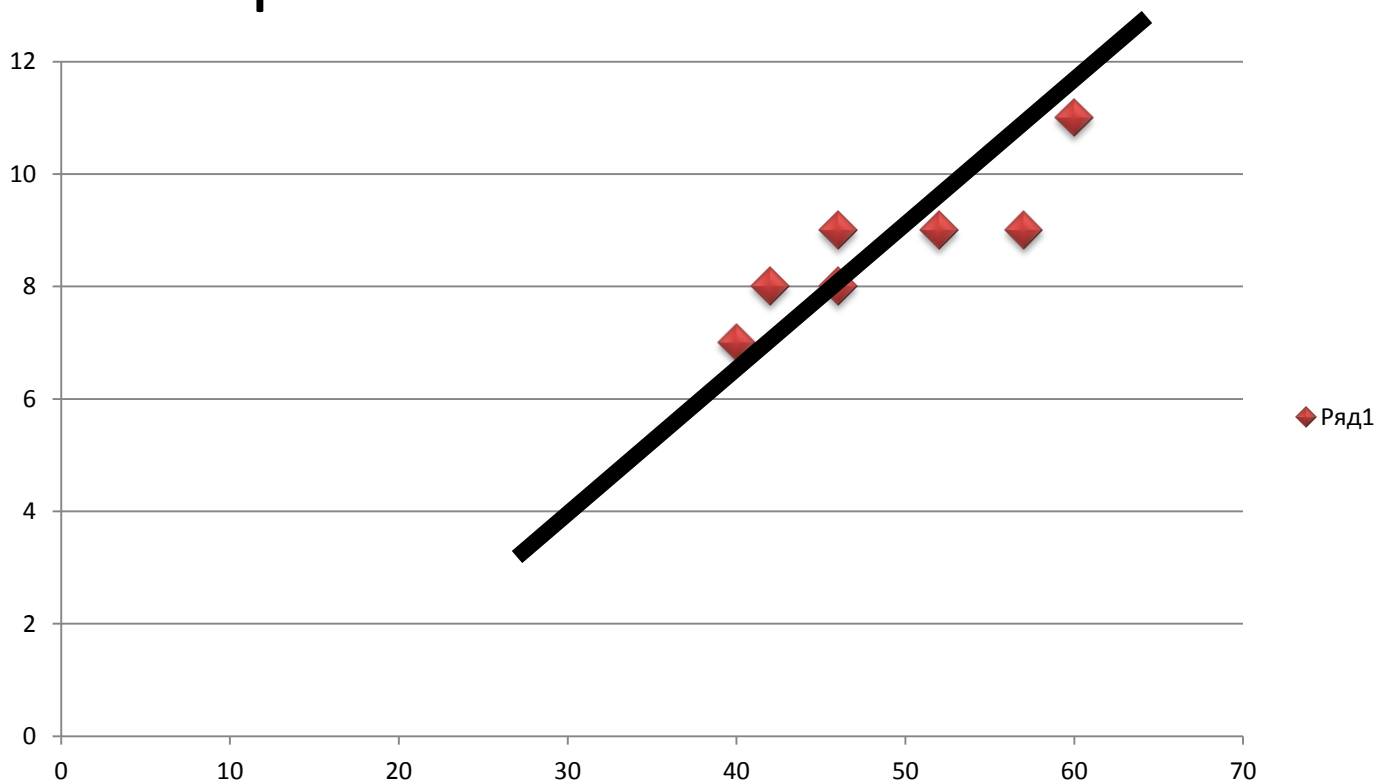
# линейная регрессия

корреляционную связь признаков  $Y$  и  $X$ , можно изобразить **точками на плоскости**, построив систему прямоугольных координат



# линейная регрессия

довольно часто корреляционная связь выглядит в виде прямой или может быть аппроксимирована прямой линией.



# линейная регрессия

линейная зависимость между переменными  $Y$  и  $X$  описывается уравнением общего вида:

$$\bar{y}_x = a + bx_1 + cx_2 + dx_3 + \dots,$$

где  $a, b, c, d, \dots$  - параметры уравнения, определяющие соотношения между аргументами  $x_1, x_2, x_3, \dots, x_m$  и функцией  $\bar{y}_x$ .

# линейная регрессия

начиная со второго уровня, т.е. когда появляется  $x_2$  и далее  $x_3, x_n$  водится понятие множественной линейной регрессии

на практике учитывают не все возможные, а лишь некоторые аргументы, в простейшем случае – всего один:

$$\bar{y}_x = a + bx$$

# линейная регрессия

$$\bar{y}_x = a + bx$$

$x$  и  $y$  коррелирующие в своей вариации величины;

$a$  – первоначальное значение  $y$  при  $x=0$ ;

$b$  – коэффициент пропорциональности, который показывает степень зависимости  $x$  от  $y$  и определяет наклон линии регрессии по отношению к осям прямоугольных координат (коэффициент регрессии)

# линейная регрессия

поскольку показатели регрессии выражают корреляционную связь **двусторонне**, то уравнение регрессии следует записывать так:

$$\bar{y}_x = a_{yx} + b_{yx} x$$

$$\bar{x}_y = a_{xy} + b_{xy} y$$



# коэффициент регрессии

**коэффициент регрессии** показывает, насколько в среднем величина одного признака  $y$  изменяется при изменении на единицу меры другого, корреляционно связанного с  $Y$  признака  $X$

$$b_{yx} = r_{xy} \sqrt{\frac{\sum (y_i - \bar{y})^2}{\sum (x_i - \bar{x})^2}} \quad b_{xy} = r_{xy} \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}}$$

# коэффициент регрессии

$$b_{yx} = r_{xy} \sqrt{\frac{\sum (y_i - \bar{y})^2}{\sum (x_i - \bar{x})^2}} \quad b_{xy} = r_{xy} \sqrt{\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N-1}}$$

$$b_{yx} = r_{yx} \frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = r_{xy} \frac{\sigma_x}{\sigma_y}$$

# коэффициент регрессии

если же коэффициент корреляции не известен, то коэффициент регрессии определяют следующим образом

$$b_{yx} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$b_{xy} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (y_i - \bar{y})^2}$$

# линейная регрессия в Excel

уравнение с одним аргументом  $\bar{y}_x = a + bx$

в пакете анализа данных Excel экспериментальные данные аппроксимируются линейным уравнением до 16 порядка:

$$\bar{y}_x = a_0 + a_1x_1 + a_2x_2 + \dots + a_{16}x_{16},$$

где  $y_x$  – зависимая переменная;

$x_1, \dots, x_{16}$  – независимые переменные;  $a_0, a_1,$

$\dots, a_{16}$  – искомые коэффициенты регрессии.

# линейная регрессия в Excel

что мы получаем при анализе линейной регрессии в Excel?

- ✓ коэффициент регрессии
- ✓ показатели регрессионной статистики (R-квадрат)
- ✓ значимость F регрессии
- ✓ график линейной регрессии

# линейная регрессия в Excel

коэффициенты регрессии в Excel:

$a_0$  - « $Y$ -пересечение»

$a_1$  - «переменная  $x_1$ »

$$\bar{y}_x = a_0 + a_1 x_1$$

# линейная регрессия в Excel

коэффициент детерминации  
(**R-квадрат**) оценивает «силу  
влияния» переменной  $x$  на  $y$ .

$R\text{-квадрат} > 0,95$  – высокая точность аппроксимации;

$0,8 < R\text{-квадрат} < 0,95$  – удовлетворительная;

$R\text{-квадрат} < 0,6$  – точность не достаточна и модель требует улучшения.

# линейная регрессия в Excel

значимость F

не должна превышать 0,05

значимость F говорит о высокой достоверности результатов и **отсутствии случайности** и наличия оправданной в нашем исследовании закономерности



# линейная регрессия в Excel

**дополнительно** оценку  
достоверности результатам  
регрессионного анализа можно  
провести с помощью критерия  
Стьюдента

**в Excel терминология:**

«стандартная ошибка» и  
«t-статистика»

# линейная регрессия в Excel

когда достоверность с помощью коэффициентов доказана можно строить регрессионную модель (уравнение линейной регрессии):

$$\bar{y}_x = a + bx$$

или  $Y = a_0 + a_1x_1$  (в Excel)

# линейная регрессия в Excel

для исключения коэффициента -  $a$   
(функция будет проходить через  
начало координат) в окне  
регрессионного анализа  
необходимо включить опцию  
«**константа-ноль**»

«Р-значения»  $> 0,05$ .

# линейная регрессия в Excel

уравнение линейной регрессии:

полученные в результате

регрессионного анализа

коэффициенты *a* и *b* мы вставляем в

уравнение линейной регрессии

$$\bar{y}_x = a + bx$$

# линейная регрессия в Excel

**теоретическая линия регрессии:**

в готовое уравнение мы

подставляем значения  $x$  и строим

график в виде линии

**эмпирическая линия регрессии:**

ломанная линия соединяющая все

реальные (эмпирические) данные

исследования

# линейная регрессия в Excel



# степень достоверности

линии регрессии и коэффициента регрессии

степень достоверности устанавливается по величине  $t$ :

$t = b / S_b$ , где

$b$  – это коэффициент регрессии,

$S_b$  – ошибка коэффициента регрессии

Степень свободы ( $df$ ) берется как  $n-2$

# степень достоверности

линии регрессии и коэффициента регрессии

ошибка коэффициента регрессии:

$$S_b = \frac{\sigma_{yx}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

с выражением **СИГМЫ** выглядит так

$$S_b = \sqrt{\frac{\sum (y_i - \bar{y})^2 - \frac{[\sum (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum (x_i - \bar{x})^2}}{(n-2) \sum (x_i - \bar{x})^2}}$$



# степень достоверности

## линии регрессии и коэффициента регрессии

Значения  $t$  при различных уровнях значимости ( $P$ )

Число степеней свободы	Уровень значимости $P$				
	0,1	0,05	0,02	0,01	0,001
1	6,31	12,7	31,82	63,66	—
2	2,92	4,30	6,97	9,93	31,60
3	2,35	3,18	4,54	5,84	12,94
4	2,13	2,78	3,75	4,60	8,61
5	2,02	2,57	3,37	4,03	6,86
6	1,94	2,45	3,14	3,71	5,96
7	1,90	2,37	3,00	3,50	5,41
8	1,86	2,31	2,90	3,36	5,04
9	1,83	2,26	2,82	3,25	4,78
10	1,81	2,23	2,76	3,17	4,59
11	1,80	2,20	2,72	3,11	4,44
12	1,78	2,18	2,68	3,06	4,32
13	1,77	2,16	2,65	3,01	4,22
14	1,76	2,15	2,63	2,98	4,14

# что же такое нулевая гипотеза?

гипотеза (нулевая):

считается, что изучаемый фактор А  
(В, С и тд.) не влияет

если гипотеза верна, то:

дисперсия  $\sigma_A^2$  равна нулю, т.е.  
вариация случайна

# что же такое нулевая гипотеза?

проводя анализ (к, р, д) мы доказываем

**несостоятельность гипотезы:**

$\sigma_A^2$  не равна 0

! но при этом входит в уровень  
значимости 0,05 или с вероятностью 0,95

достоверность же дисперсии  $\sigma_A^2$   
устанавливается путем деления на  
свою ошибку

# линейный регрессионный анализ

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вставить Буфер обмена Шрифт Выравнивание

Общий Вставить  $\Sigma$

A16 fx

	A	B	C	D
1	N	Время, мин	Толщина рубца, мм	
2	1	2,4	17	
3	2	0,6	3	
4	3	1,7	12	
5	4	1	5	
6	5	1,6	8	
7	6	1,5	9	
8	7	1,8	14	
9	8	3	20	
10				
11	Имеются результаты проводившиеся у 8 больных эффективной криодеструкции			
12	кожных рубцов различной толщины			
13				
14	Необходимо рассмотреть возможность на основании этих данных определять			
15	предполагаемое время криодеструкции			
16				

Лист1 Лист2 Лист3

Готово 100%

Линейный регрессионный анализ заключается в подборе графика и его уравнения для набора наблюдений

Регрессия используется для анализа воздействия на отдельную зависимую переменную значений одной или более независимых переменных

# линейный регрессионный анализ

В пакете анализа данных Excel экспериментальные данные аппроксимируются линейным уравнением до 16 порядка:

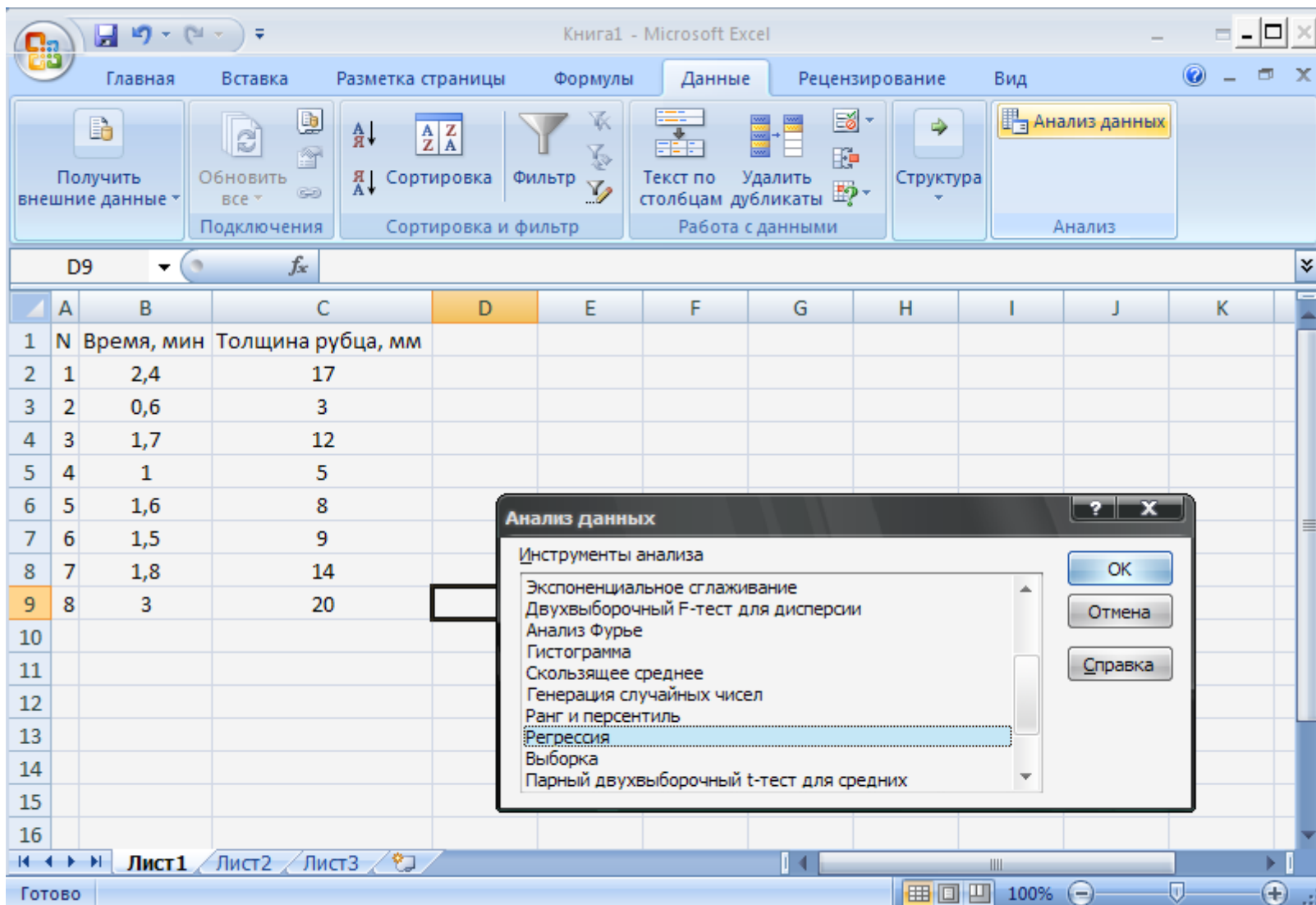
$$Y = a_0 + a_1 * x_1 + a_2 * x_2 + \dots + a_{16} * x_{16},$$

где  $Y$  – зависимая переменная;  $x_1, \dots, x_{16}$  – независимые переменные;  $a_0, a_1, \dots, a_{16}$  – искомые коэффициенты регрессии.

Далее показано как получить эти коэффициенты регрессии



# линейный регрессионный анализ



Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы **Данные** Рецензирование Вид

Получить внешние данные Обновить все Подключения

Сортировка и фильтр

Работа с данными

Анализ данных

Анализ

D9 fx

	A	B	C	D	E	F	G	H	I	J	K
1	N	Время, мин	Толщина рубца, мм								
2	1	2,4	17								
3	2	0,6	3								
4	3	1,7	12								
5	4	1	5								
6	5	1,6	8								
7	6	1,5	9								
8	7	1,8	14								
9	8	3	20								
10											
11											
12											
13											
14											
15											
16											

Анализ данных

Инструменты анализа

- Экспоненциальное сглаживание
- Двухвыборочный F-тест для дисперсии
- Анализ Фурье
- Гистограмма
- Скользящее среднее
- Генерация случайных чисел
- Ранг и перцентиль
- Регрессия**
- Выборка
- Парный двухвыборочный t-тест для средних

OK Отмена Справка

Лист1 Лист2 Лист3

Готово 100%

# линейный регрессионный анализ

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы **Данные** Рецензирование Вид

Получить внешние данные Обновить все Подключения Сортировка и фильтр Сортировка Фильтр Работа с данными Текст по столбцам Удалить дубликаты Структура Анализ данных

	A	B	C	D
1	N	Время, мин	Толщина рубца, мм	
2	1	2,4	17	
3	2	0,6	3	
4	3	1,7	12	
5	4	1	5	
6	5	1,6	8	
7	6	1,5	9	
8	7	1,8	14	
9	8	3	20	
10				
11				
12				
13				
14				
15				
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

Лист1 Лист2 Лист3

Ввод 100%



# линейный регрессионный анализ

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вставить Буфер обмена Шрифт Выравнивание Число Стили Ячейки Редактирование

Д1 *f<sub>x</sub>* ВЫВОД ИТОГОВ

	A	B	C	D	E	F	G
3	2	0,6	3	<i>Регрессионная статистика</i>			
4	3	1,7	12	Множественный R	0,972338443		
5	4	1	5	R-квadrat	0,945442047		
6	5	1,6	8	Нормированный R-квadrat	0,936349055		
7	6	1,5	9	Стандартная ошибка	0,189278601		
8	7	1,8	14	Наблюдения	8		
9	8	3	20				
10				<i>Дисперсионный анализ</i>			
11					<i>df</i>	<i>SS</i>	<i>MS</i>
12				Регрессия	1	3,725041667	3,725
13				Остаток	6	0,214958333	0,03583
14				Итого	7	3,94	
15							
16					<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>
17				Y-пересечение	0,329583333	0,150135837	2,195
18				Переменная X 1	0,124583333	0,012217881	10,195

Лист1 Лист2 Лист3

Готово Среднее: 5,017902256 Количество: 57 Сумма: 155,5549699 100%

# линейный регрессионный анализ

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы **Данные** Рецензирование Вид Разработчик

Получить внешние данные Обновить все Подключения

Сортировка Фильтр Сортировка и фильтр

Очистить Применить повторно Дополнительно

Текст по столбцам Удалить дубликаты Работа с данными

Структура Анализ данных Анализ

C1 *f<sub>x</sub>* ВЫВОД ИТОГОВ

	B	C	D	E	F	G	H
4	5	Множественный R	0,972338443				
5	8	R-квадрат	0,945442047				
6	9	Нормированный R-квадрат	0,936349055				
7	14	Стандартная ошибка	0,189278601				
8	20	Наблюдения	8				
9							
10		Дисперсионный анализ					
11			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12		Регрессия	1	3,725041667	3,725041667	103,974801	5,18222E-05
13		Остаток	6	0,214958333	0,035826389		
14		Итого	7	3,94			
15							
16			<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
17		Y-пересечение	0,329583333	0,150135837	2,195234256	0,070563987	-0,037785826
18		Переменная X 1	0,124583333	0,012217881	10,19680349	5,18222E-05	0,094687255
19							
20							
21							

Лист1 Лист2 Лист3

Готово Среднее: 5,017902256 Количество: 57 Сумма: 155,5549699 100%

# линейный регрессионный анализ

Интерпретация результатов  
(коэффициенты регрессии):

	<i>Коэффициенты</i>
Y-пересечение	0,329583333
Переменная X 1	0,124583333

Y-пересечение –  $a_0$ ;

Переменная  $x_1$  –  $a_1$ ;

# линейный регрессионный анализ

## Интерпретация результатов:

<i>Регрессионная статистика</i>	
Множественный R	0,972338443
R-квадрат	0,945442047
Нормированный R-квадрат	0,936349055
Стандартная ошибка	0,189278601
Наблюдения	8

R-квадрат  $> 0,95$  – высокая точность аппроксимации;

$0,8 < \text{R-квадрат} < 0,95$  – удовлетворительная;

R-квадрат  $< 0,6$  – точность не достаточна и модель требует улучшения.

# линейный регрессионный анализ

Интерпретация результатов:

Значимость F
5,18222E-05

Значимость F =  $5,18222E-05 = 5,18222 * 10^{-5}$ , что меньше 0,05

это говорит о высокой достоверности результатов, т.к. значение случайной дисперсии низко по сравнению с факториальной

# линейный регрессионный анализ

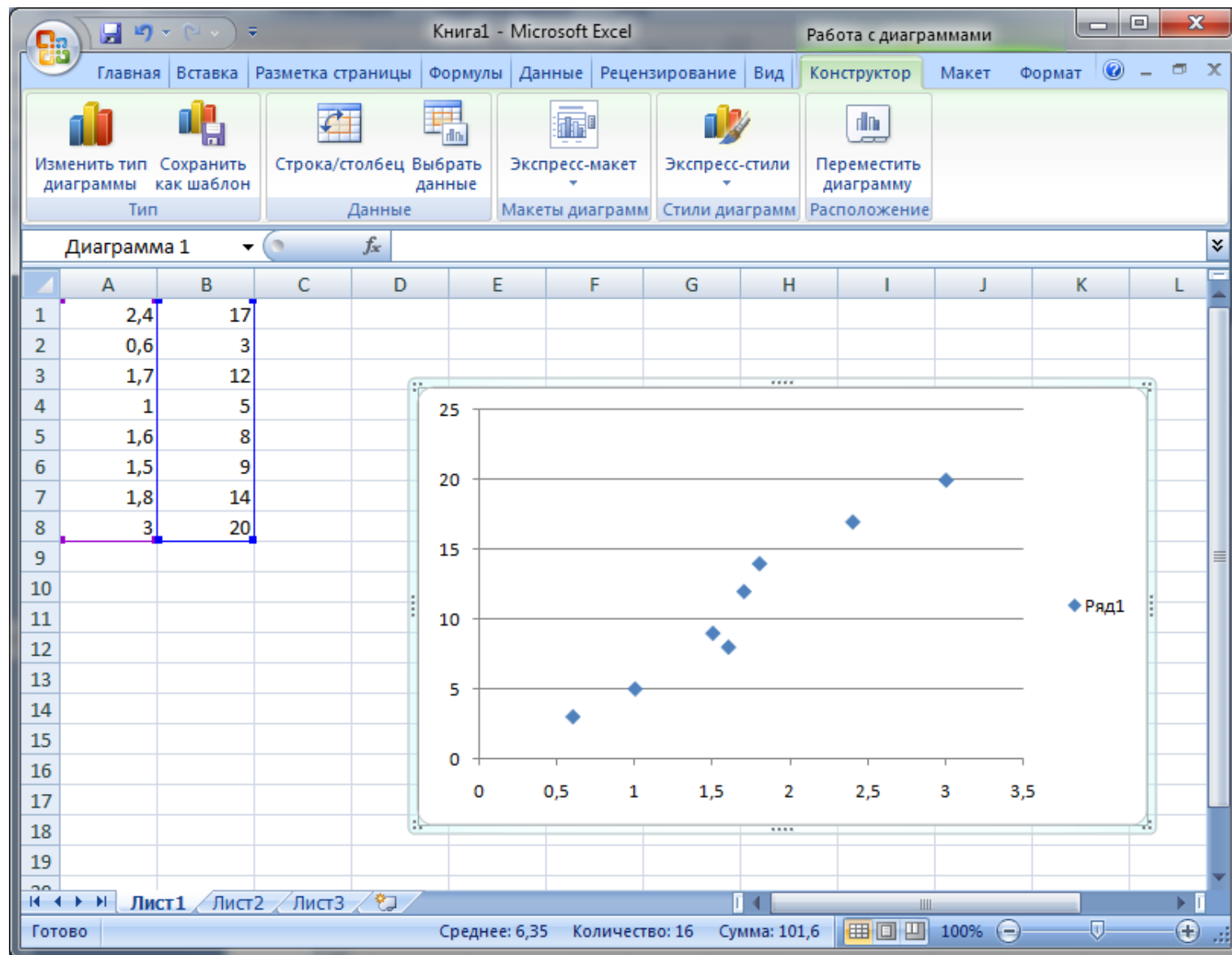
В итоге, время криодеструкции в мин =  
0,33 (Y-пересечение) +  
0,125 (переменная X1)\*  
(толщина рубца в мм)

$$Y = a_0 + a_1 * x_1,$$

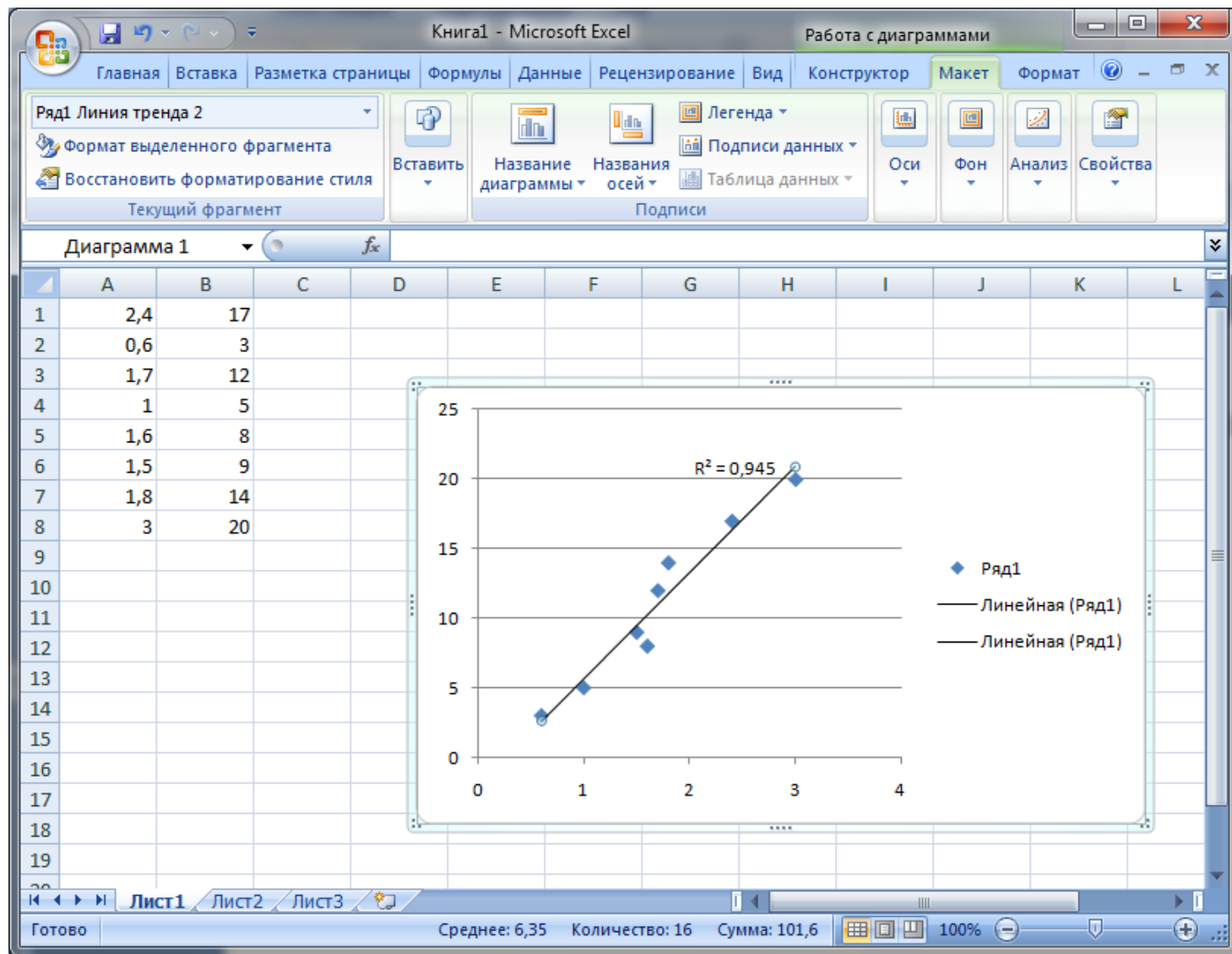
$$Y = 0,33 + 0,125 * 17 = 2,455$$

(например у больного №1)

# линейный регрессионный анализ



# линейный регрессионный анализ





# криволинейная регрессия

связь биологических признаков не бывает линейной и биологические процессы часто развиваются по кривой-  
уравнения кривых линий

! но прямой расчет уравнений кривых по исходным данным невозможен

# криволинейная регрессия

**метод наименьших квадратов:**

минимизируется сумма квадратов отклонений реально наблюдаемых из расчетов с помощью линейного уравнения

$$\underline{y}_x = a_0 + a_1x_1 + a_2x_2 + \dots + a_Nx_N$$

# криволинейная регрессия

«выпрямление» зависимости  $y$  от  $x$

вычисление линейного уравнения

уравнение криволинейной связи

# криволинейная регрессия

## Исследование криволинейной регрессионной зависимости в Excel:

1. Заполнение исходными данными значений  $x$  (ось абсцисс) и  $y$  (ось ординат);
2. По исходным данным строим точечную диаграмму;
3. Добавляем линию тренда и выбираем наиболее оптимальную модель в «формате линии тренда» (макс. значение  $R^2$  – коэффиц. детерминации);

# криволинейная регрессия

## Исследование криволинейной регрессионной зависимости в Excel:

4. Преобразование («выпрямление») исходных данных, если выбранная модель отличается от линейной (логарифм, степень и пр.);
5. Полученные преобразованные данные подвергаются регрессионному анализу;

# криволинейная регрессия

## Исследование криволинейной регрессионной зависимости в Excel:

6. Проводится оценка значимости коэффициентов регрессии, построение уравнения с помощью этих коэффициентов, делается вывод об адекватности модели путем сравнения R-квадрата при линейной регрессии и регрессии по выбранной модели.

# полиномиальная регрессия

Полином, или многочлен,  
выражение вида:

$$\bar{y}_x = a_0 + a_1 x_1 + a_2 x_2^2 + \dots + a_N x_N^N$$

любую непрерывную функцию  
можно аппроксимировать  
(приблизительно описать)  
многочленом

# полиномиальная регрессия

Полином *первого* порядка – это прямая, описываемая уравнением  $\bar{y}_x = a_0 + a_1 x_1$ .

на переменные влияет лишь *один процесс*, его можно описать с помощью линейной регрессии



# полиномиальная регрессия

Полином *второго* порядка  
есть парабола:

$$\bar{y}_x = a_0 + a_1 x_1 + a_2 x_2^2$$

график имеет один изгиб, что  
позволяет описывать влияние на  
переменную  $y$  *двух* процессов

# полиномиальная регрессия

Полином *третьего* порядка:

$$\bar{y}_x = a_0 + a_1x_1 + a_2x_2^2 + a_3x_3^3$$

график представляет собой кривую с двумя изгибами, отражающий воздействие на переменную **нескольких** процессов

# полиномиальная регрессия

## полиномиальная регрессия в Excel:

1. Заполнение исходными данными значений  $x$  и  $y$ ;
2. По исходным данным строим точечную диаграмму;
3. Оцениваем полученную точечную диаграмму (в виде параболы, то значит полином второго порядка и т.д.);

# полиномиальная регрессия

## полиномиальная регрессия в Excel:

4. Добавляем линию тренда и выбираем полиномиальную модель и задаем порядок полинома (Степень), равный 2.
5. Ставим флажки в пунктах «показывать уравнение на диаграмме» и «поместить на диаграмму величину достоверности аппроксимации».

# множественная линейная регрессия

эффект влияния **нескольких признаков** или факторов на переменную изучают с помощью множественной регрессии

**Цель анализа:** установление зависимости среднего значения переменной  $y$  от нескольких переменных  $x_1, x_2, x_n$

Уравнение множественной регрессии:

$$\bar{y}_x = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

# множественная линейная регрессия

при необходимости в уравнение множественной регрессии можно добавить **квадратичные члены**  $a_i x_i^2$  или **трансформированные переменные**  $a_i X_i^2$

# множественная линейная регрессия

в пакете анализа данных Excel  
экспериментальные данные  
аппроксимируются линейным  
уравнением **до 16 порядка**:

$$\bar{y}_x = a_0 + a_1x_1 + a_2x_2 + \dots + a_{16}x_{16}$$

# множественная линейная регрессия

## правило множественного регрессионного анализа в Excel:

в графу «входной интервал  $Y$ » вводим ссылку на один столбец со значениями зависимой переменной  $y$ , а в графу «входной интервал  $X$ » вводим ссылку на блок ячеек с несколькими переменными  $x$ .



# множественная линейная регрессия

первое, если **R-квадрат  $> 0,6$**

то необходимо провести **второе**

(пошаговый регрессионный анализ,

т.е. анализ коэффициентов регрессии

на **P-значение**):

если значение коэффициента на пересечении

P-значения **меньше  $0,05$** , то нужно исключать

целиком в анализе данную графу с переменными  $x$ .

При этом **исключать нужно в первую очередь**

**переменную с самым низким P-значением.**

# множественная линейная регрессия

периодически исключая фактор  
необходимо заново проводить  
регрессионный анализ с уже  
уменьшенными на одну графу данными

для необходимости исключения  
свободного члена  $a_0$  (Y-пересечение)  
нужно поставить флажок  
«**Константа-ноль**»

# множественная линейная регрессия

для интерпретации результатов важно помнить, что **знак ( + )** перед коэффициентом говорит о **прямой** связи между  $y$  и  $x$  (параллельный рост значений), **знак ( - )** свидетельствует об **обратной** зависимости (при возрастании  $x$  и  $y$  уменьшается).

# связь между коэффициентами регрессии и корреляции

сравним формулы

$$b_{yx} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

или

$$b_{xy} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (y_i - \bar{y})^2}$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

# связь между коэффициентами регрессии и корреляции

эта связь выражается равенством

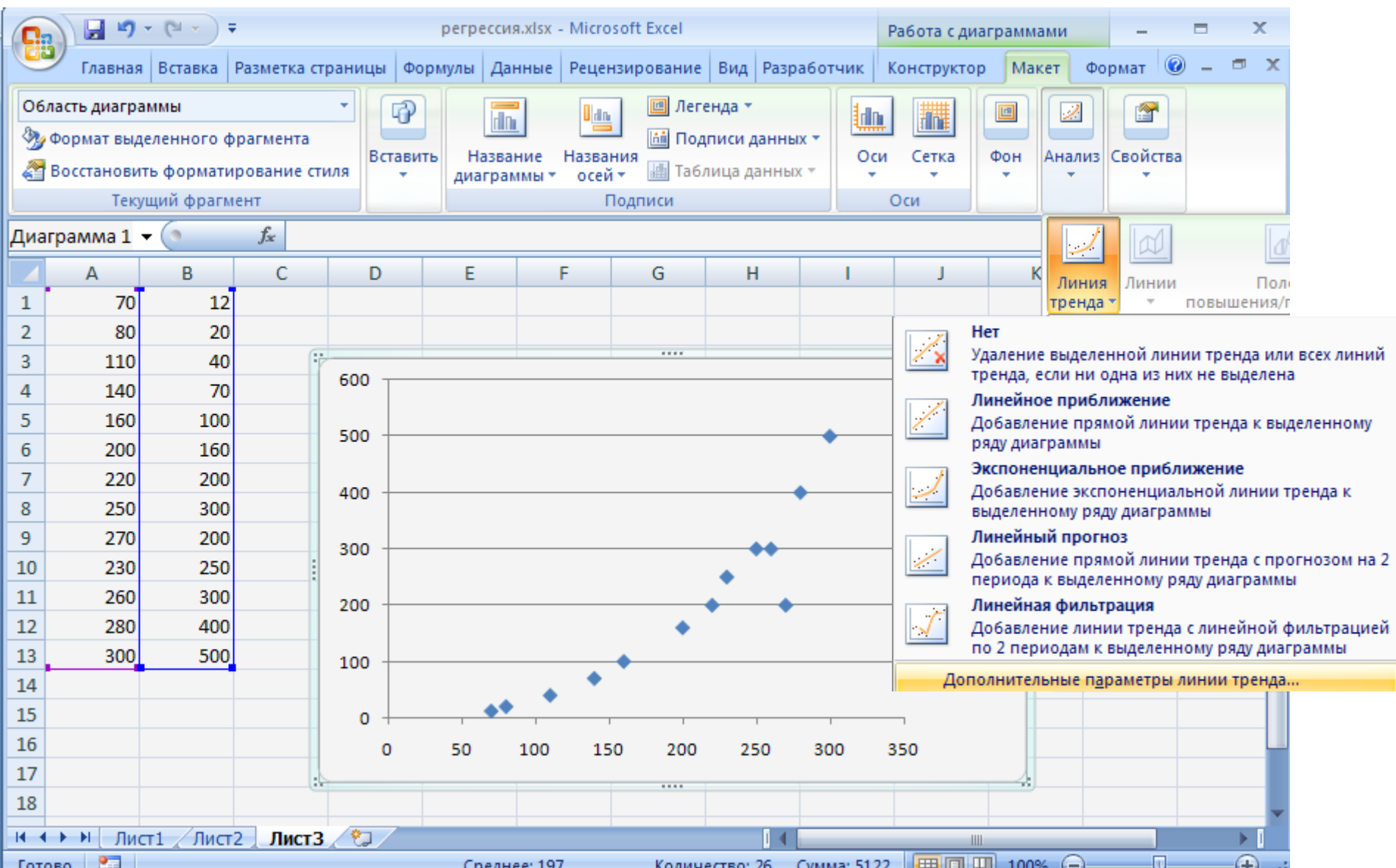
$$r_{xy}^2 = b_{yx} b_{xy} \quad \text{или} \quad r_{xy} = \sqrt{b_{yx} b_{xy}}$$

эта формула позволяет, **во-первых**, по известным значениям коэффициентов регрессии  $b_{yx}$  и  $b_{xy}$  определять коэффициент корреляции  $r_{xy}$ , а **во-вторых**, проверять правильность расчетов этого показателя корреляционной связи между варьирующими признаками  $X$  и  $Y$ .

# обзор практической части в Excel



# криволинейная регрессия





# криволинейная регрессия

Формат линии тренда

Параметры линии тренда

Построение линии тренда (аппроксимация и сглаживание)

- Экспоненциальная
- Линейная
- Логарифмическая
- Полиномиальная Степень: 2
- Степенная
- Линейная фильтрация Точки: 2

Название аппроксимирующей (сглаженной) кривой

- автоматическое: Степенная (Ряд 1)
- другое:

Прогноз

вперед на: 0,0 периодов

назад на: 0,0 периодов

пересечение кривой с осью Y в точке: 0,0

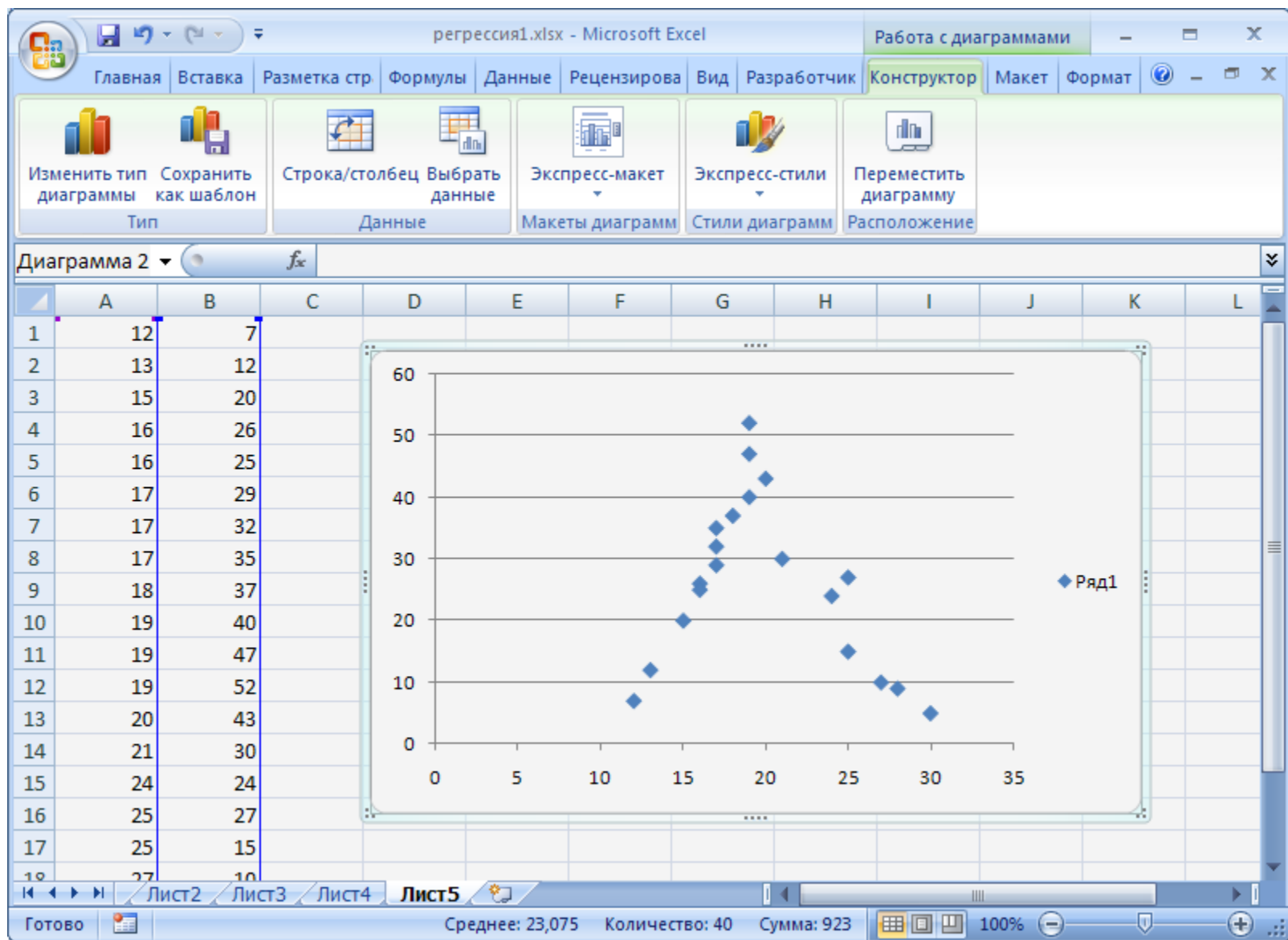
- показывать уравнение на диаграмме
- поместить на диаграмму величину достоверности аппроксимации ( $R^2$ )

Закреть

Выбор модели с максимальным значением  $R^2$  (коэффициентом детерминации)

	A	B	C	D	E
1	70	12			
2	80	20			
3	110	40			
4	140	70			
5	160	100			
6	200	160			
7	220	200			
8	250	300			
9	270	200			
10	230	250			
11	260	300			
12	280	400			
13	300	500			
14					
15					
16					
17					
18					

# полиномиальная регрессия



# полиномиальная регрессия

Формат линии тренда

Параметры линии тренда

Цвет линии  
Тип линии  
Тень

Построение линии тренда (аппроксимация и сглаживание)

- Экспоненциальная
- Линейная
- Логарифмическая
- Полиномиальная Степень: 2
- Степенная
- Линейная фильтрация Точки: 2

Название аппроксимирующей (сглаженной) кривой

- автоматическое: Полиномиальная (Ряд 1)
- другое:

Прогноз

вперед на: 0,0 периодов  
назад на: 0,0 периодов

пересечение кривой с осью Y в точке: 0,0

показывать уравнение на диаграмме  
 поместить на диаграмму величину достоверности аппроксимации ( $R^2$ )

Заккрыть

	A
1	12
2	13
3	15
4	16
5	16
6	17
7	17
8	17
9	18
10	19
11	19
12	19
13	20
14	21
15	24
16	25
17	25
18	27

Среднее: 23,075 Количество: 40 Сумма: 923 100%

# множественная линейная регрессия

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вставить Буфер обмена Шрифт Выравнивание Число Стили Ячейки Редактирование

Общий % 000

	A	B	C	D	E	F	G	H	I	J	K	L
1	Y	X1	X1									
2	2200	3,2	22									
3	1600	3,5	25									
4	700	4,3	30									
5	400	4	40									
6	1100	3,8	30									
7	800	3,6	39									
8	700	4,2	30									
9	1100	3,3	39									
10	1100	4,1	26									
11	1800	3,4	23									
12												
13												
14												
15												
16												

Лист1 Лист2 Лист3

Готово 100%





# множественная линейная регрессия

The screenshot shows the Microsoft Excel interface with the 'Регрессия' (Regression) dialog box open. The dialog box is titled 'Регрессия' and has a red box around the 'Выходной интервал:' (Output Range) option, which is selected with a radio button. The 'Выходной интервал:' field contains the cell reference '\$D\$1'. Other options in the dialog include 'Входной интервал Y:' (\$A\$2:\$A\$11), 'Входной интервал X:' (\$B\$2:\$C\$11), 'Метки' (Labels), 'Константа - ноль' (Constant zero), 'Уровень надежности:' (Confidence level) set to 95%, 'Новый рабочий лист:' (New worksheet), 'Новая рабочая книга' (New workbook), 'Остатки' (Residuals) with options for 'Остатки' (Residuals), 'График остатков' (Residuals chart), 'Стандартизованные остатки' (Standardized residuals), and 'График подбора' (Best-fit line chart), and 'Нормальная вероятность' (Normal distribution) with 'График нормальной вероятности' (Normal distribution chart).

	A	B	C	D	E
1	Y	X1	X1		
2	2200	3,2	22		
3	1600	3,5	25		
4	700	4,3	30		
5	400	4	40		
6	1100	3,8	30		
7	800	3,6	39		
8	700	4,2	30		
9	1100	3,3	39		
10	1100	4,1	26		
11	1800	3,4	23		
12					
13					
14					
15					
16					

# множественная линейная регрессия

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вставить Буфер обмена Шрифт Выравнивание Число Стили Ячейки Редактирование

Вывод итогов

	A	B	C	D	E	F	G
1	Y	X1	X1	Вывод итогов			
2	2200	3,2	22				
3	1600	3,5	25	<i>Регрессионная статистика</i>			
4	700	4,3	30	Множественный R	0,995673617		
5	400	4	40	R-квадрат	0,991365951		
6	1100	3,8	30	Нормированный R-квадрат	0,98889908		
7	800	3,6	39	Стандартная ошибка	59,02927654		
8	700	4,2	30	Наблюдения	10		
9	1100	3,3	39				
10	1100	4,1	26	<i>Дисперсионный анализ</i>			
11	1800	3,4	23		<i>df</i>	<i>SS</i>	<i>MS</i>
12				Регрессия	2	2800608,812	1400304,40
13				Остаток	7	24391,18842	3484,45548
14				Итого	9	2825000	
15							
16				<i>Коэффициенты Стандартная ошибка t-статистика</i>			

Лист1 Лист2 Лист3

Готово Среднее: 181560,9651 Количество: 66 Сумма: 7080877,64 100%



# множественная линейная регрессия

Книга1 - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Вставить Буфер обмена Шрифт Выравнивание Число Стили Ячейки Редактирование

D1  $f_x$  ВЫВОД ИТОГОВ

	D	E	F	G	H	I
7	Стандартная ошибка	59,02927654				
8	Наблюдения	10				
9						
10	Дисперсионный анализ					
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12	Регрессия	2	2800608,812	1400304,406	401,8718019	5,98068E-08
13	Остаток	7	24391,18842	3484,455489		
14	Итого	9	2825000			
15						
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
17	Y-пересечение	6244,294904	196,4001692	31,79373485	7,87154E-09	5779,882301
18	Переменная X 1	-919,4176401	50,46127644	-18,22026126	3,71273E-07	-1038,739598
19	Переменная X 2	-54,46292534	2,937189131	-18,54253264	3,2912E-07	-61,40827395
20						
21						
22						

Лист1 Лист2 Лист3

Готово Среднее: 181560,9651 Количество: 66 Сумма: 7080877,64 100%

# множественная линейная регрессия

## Интерпретация результатов:

<i>Регрессионная статистика</i>	
Множественный R	0,995673617
R-квадрат	0,991365951
Нормированный R-квадрат	0,98889908
Стандартная ошибка	59,02927654
Наблюдения	10

R-квадрат  $> 0,95$  – высокая точность результата

# множественная линейная регрессия

В итоге результат исследования =  
6244 (Y-пересечение) –  
919 (переменная X1) –  
54,5 (переменная X2)\* (нужную величину)

$Y = a_0 + a_1 * X_1 + a_2 * X_2 + \dots + a_{16} * X_{16}$ ,  
где Y – зависимая переменная; X1, ..., X16 –  
независимые переменные; a0, a1, ..., a16 –  
искомые коэффициенты регрессии.