Решение вычислительных задач на компьютере

Решение вычислительных задач на компьютере

§ 69. Точность вычислений

Погрешности измерений

«Недостатки математического образования с наибольшей отчетливостью проявляются в чрезмерной точности численных расчетов».
Карл Фридрих Гаусс.

Погрешность (ошибка) – отклонение измеренного или вычисленного значения от истинного значения.

цена деления 0,1 см

измерено	фактически
8,2 см	**8,15 ... 8,25** см
7,8 см	**7,75 ... 7,85** см

Толщина дна:

вычислено	фактически
0,4 см	**0,3 ... 0,5** см

0,4 ± 0,1 см

Погрешности измерений

абсолютная погрешность Δx

0,4 ± 0,1 см

Можно ли оценить точность измерений?

Относительная погрешность:

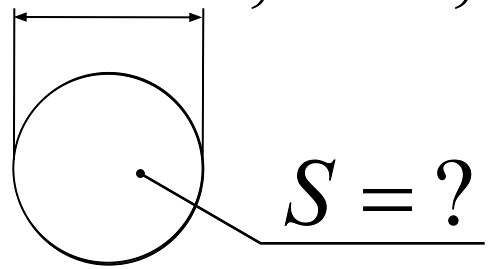
$$\delta_x = \frac{\Delta x}{x^*}$$

измеренное значение

$$\delta_x = \frac{0{,}1}{0{,}4} = 0{,}25 = 25\%$$

Погрешности вычислений

$$D = 1{,}2 \pm 0{,}1 \text{ см}$$



$$S = \frac{\pi \cdot D^2}{4} = 1{,}13097335529232556584655161798 06\ldots$$

$$S_{\min} = \frac{\pi \cdot 1{,}1^2}{4} = 0{,}950\ldots \qquad S_{\max} = \frac{\pi \cdot 1{,}3^2}{4} = 1{,}327\ldots$$

$S = 1{,}1 \pm 0{,}2$ с
м

$$\delta_x = \frac{0{,}2}{1{,}1} \cdot 100\% \approx 18\%$$

Все практические расчеты выполняются неточно. Погрешность результата вычислений определяется погрешностью исходных данных.

Погрешности вычислений

$$x = \frac{a}{b} - \frac{c}{d}$$

$$a = 1000 \pm 0{,}001; \quad b = 0{,}002 \pm 0{,}001;$$
$$c = 1000 \pm 0{,}001; \quad d = 0{,}003 \pm 0{,}001.$$

$$\delta_a = \delta_c = \frac{0{,}001}{1000} = 0{,}01\%$$

$$\delta_b = \frac{0{,}001}{0{,}002} = 50\% \qquad \delta_b = \frac{0{,}001}{0{,}003} = 33\%$$

неточные числа в знаменателе

$$x = \frac{1000}{0{,}002} - \frac{1000}{0{,}003} = 166667$$

$$x_{\max} = \frac{1000}{0{,}001} - \frac{1000}{0{,}004} = 750000$$

$$x_{\min} = \frac{1000}{0{,}003} - \frac{1000}{0{,}002} = -166667$$

$$\delta_x = \frac{750000 - 166667}{166667} \approx 352\%$$

Метод **вычислительно неустойчив**: малые погрешности в исходных данных могут привести к большим погрешностям в решении.

Источники погрешностей

- неточность **исходных данных**
- неточность записи **вещественных чисел** в двоичном коде конечной длины
- погрешности приближенного вычисления некоторых стандартных **функций** (*sin*, *cos*, …)
- накопление погрешностей при **арифметических действиях** с неточными данными
- погрешность **метода**

Решение вычислительных задач на компьютере

§ 70. Решение уравнений

Методы решения уравнений

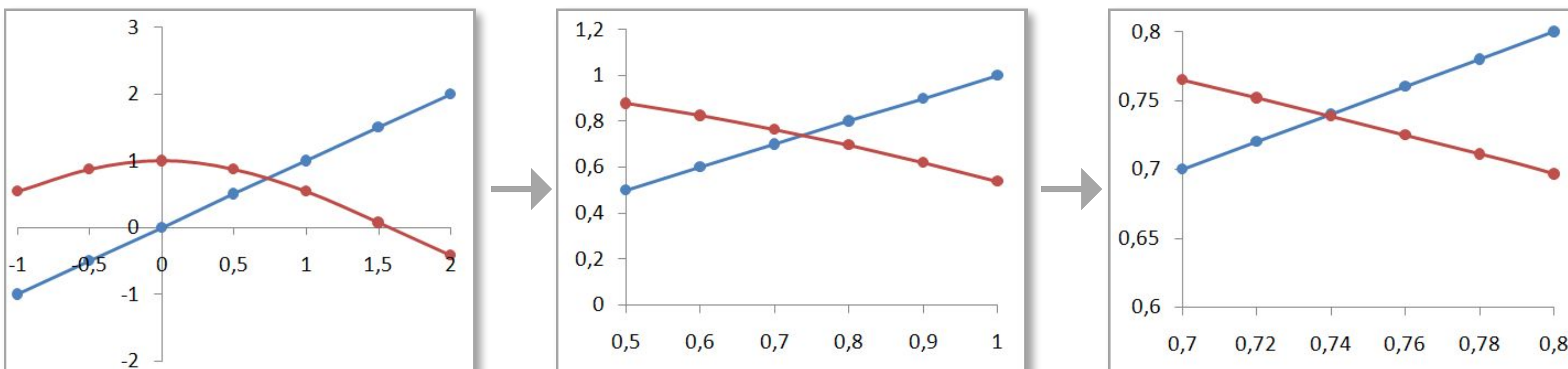
Точные (аналитические) методы:

$$ax + b = 1, \quad a \neq 0 \quad \Rightarrow \quad x = \frac{1-b}{a}$$

$$x = \cos x$$

Как решать?

Графический метод:



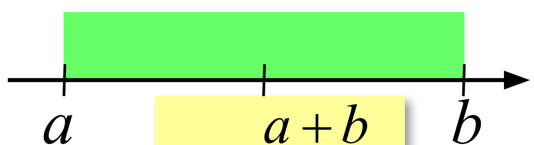
Можно поручить такой поиск компьютеру!

Можно ли получить точное решение?

Приближённые методы

Сжатие отрезка:

1) выбрать начальный отрезок $[a_0, b_0]$ (одно решение!)
2) уточнить решение с помощью некоторого алгоритма: ⇒ $[a, b]$
3) повторять шаг 2, пока длина отрезка $[a, b]$ не станет достаточно мала



$$x = \frac{a+b}{2}$$

Что лучше выбрать в качестве решения?

Как оценить ошибку?

$$\left| x - x^* \right| \le \frac{b-a}{2}$$

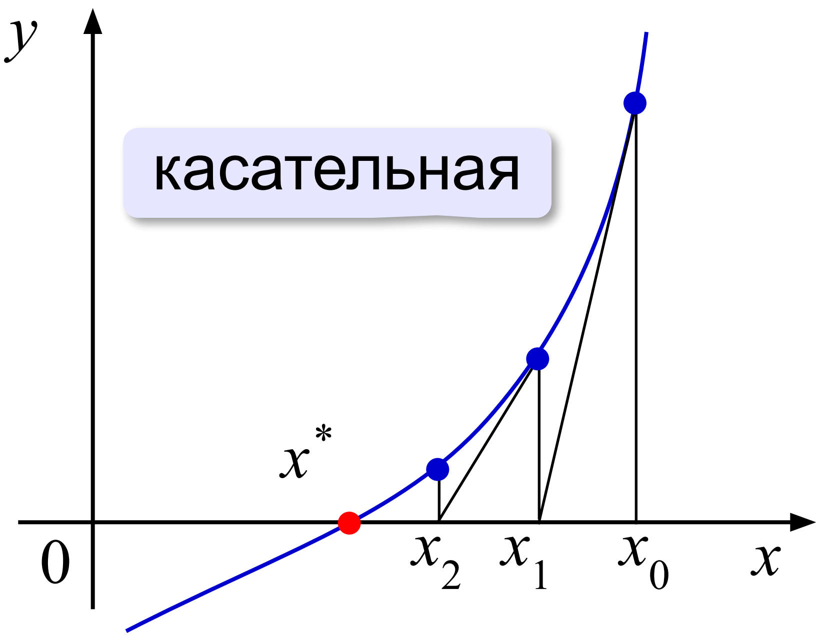
Завершение работы: $b - a \le 2\varepsilon$

допустимая ошибка

Приближенные методы

По одной точке:

1) выбрать начальное приближение x_0
2) уточнить решение с помощью некоторого алгоритма: $\Rightarrow x$
3) повторять шаг 2, пока два последовательных приближения не будут отличаться достаточно мало



Завершение работы:

$$\left| x_i - x_{i-1} \right| \le \varepsilon$$

метод Ньютона (метод касательных)

Приближенные методы

Итерационные методы (лат. *iteratio – повторение*) – основаны на многократном выполнении одинаковых шагов, каждый из которых уточняет решение.

$$x_{k+1} = f(x_k)$$

следующее приближение

предыдущее приближение

- дают какое-то решение, если точное неизвестно
- могут давать меньшие ошибки, чем вычисления по точным формулам

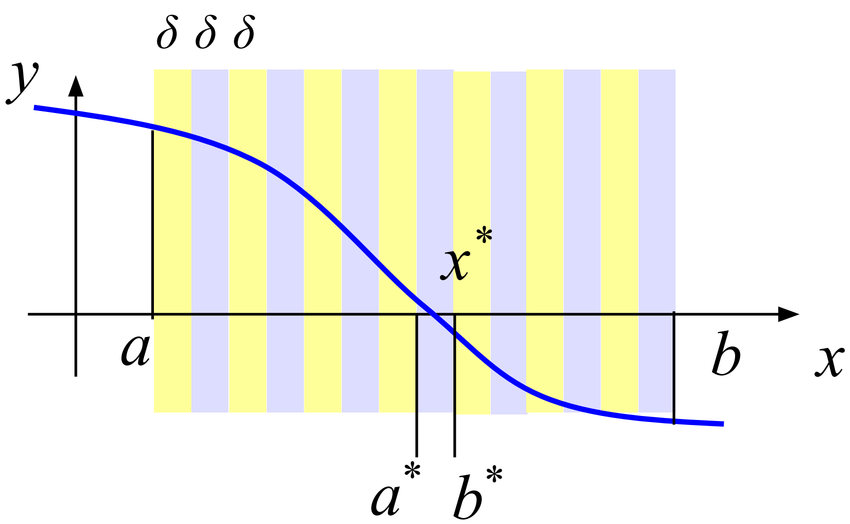
- решение приближенное: $x \boxed{\approx} 1,23345$
- ответ – число (зависимость от параметра?)
- большой объем вычислений
- не всегда просто оценить погрешность

Метод перебора

$$f(x) = 0$$

$$x = \cos x \quad \Rightarrow \quad x - \cos x = 0$$

Задача. Найти решение уравнения справа от точки $x = a$ с точностью ε.



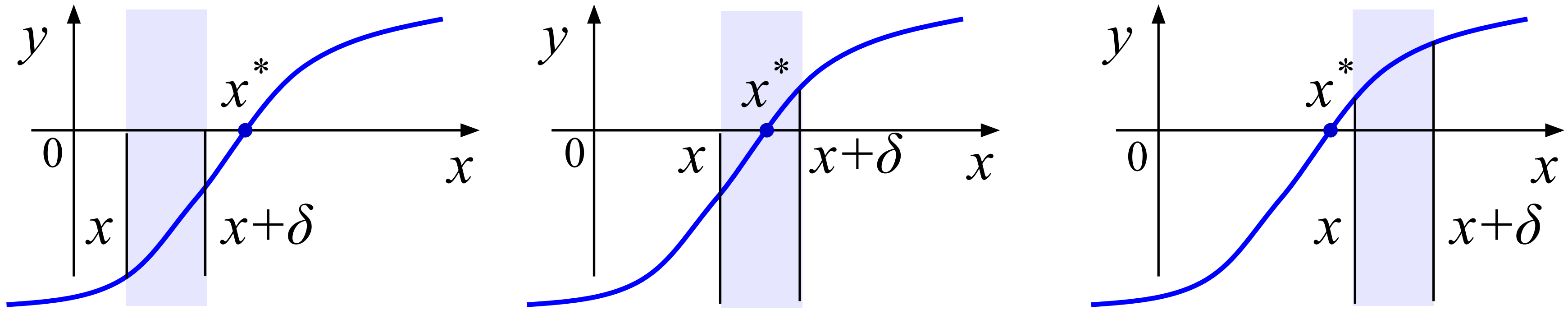
Алгоритм:

1) разбить отрезок $[a, b]$ на полосы шириной $\delta = 2\varepsilon$
2) найти полосу $[a^*, b^*]$, в которой находится x^*
3) решение:

$$x^* \approx \frac{a^* + b^*}{2}$$

Есть ли решение на $[x, x+\delta]$?

нет решения | есть решение! | нет решения



$f(x) < 0$, $f(x+\delta) < 0$

$f(x) < 0$, $f(x+\delta) > 0$

$f(x) > 0$, $f(x+\delta) > 0$

? В чём отличие?

$$f(x)f(x+\delta) \leq 0$$

! Если *непрерывная* функция $f(x)$ имеет разные знаки на концах интервала $[a, b]$, то в некоторой точке x^* внутри $[a, b]$ она равна 0, то есть $f(x^*) = 0$!

Метод перебора ($a = 0$)

```
алг Перебор
нач
  вещ eps, x, delta
  eps:= 0.001
  x:= 0   | x:= a
  delta:= 2*eps
  нц пока f(x)*f(x+delta) > 0
    x:= x + delta
  кц
  вывод 'x = ', x+eps
кон
```

Когда остановится?

Зацикливание?

```
алг вещ f( вещ x )
нач
  знач:= x - cos(x)
кон
```

Метод перебора ($a = 0$)

```
const eps = 0.001;
var x, delta: real;
 function f(x: real):real;
 begin
   f:= x - cos(x)
 end;
begin
  x:= 0; {x:= a;}
  delta:= 2*eps;
  while f(x)*f(x+delta) > 0 do
    x:= x + delta;
  writeln('x = ',(x+eps):6:3)
end.
```

Когда остановится?

Зацикливание?

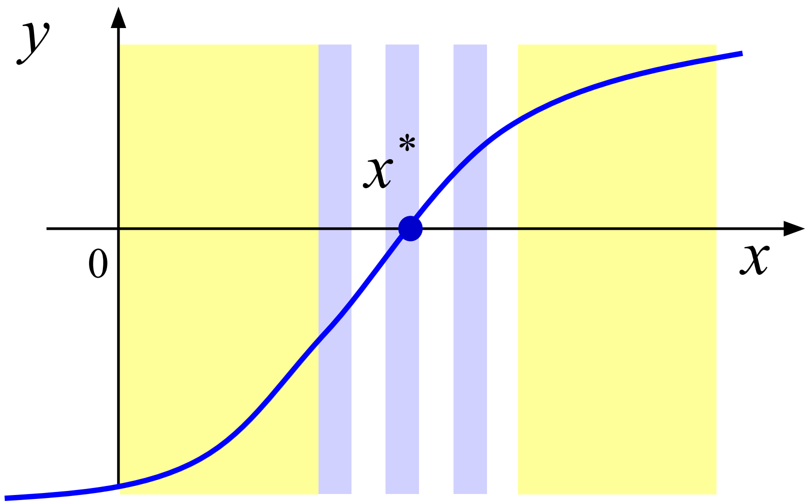
Метод перебора

- **простота**
- можно получить решение с **любой заданной точностью**

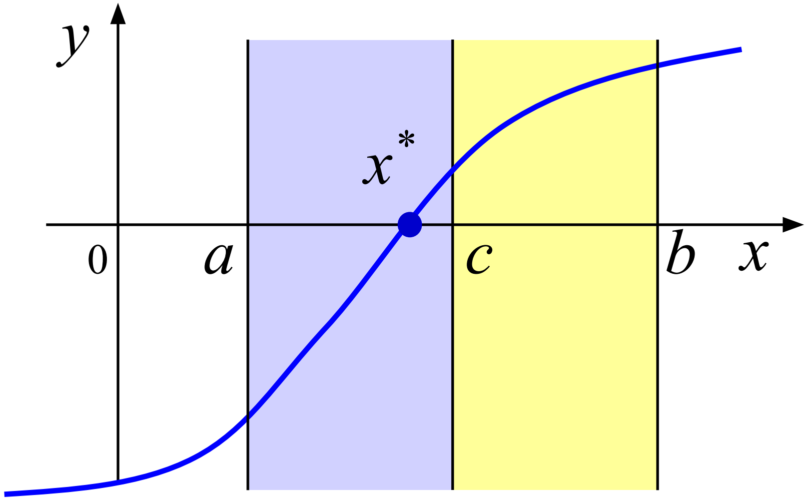
- большой **объем вычислений**

Усовершенствованный перебор:

1) **отделение корней** – перебор с большим шагом
2) **уточнение корней** – перебор с шагом 2ε



Метод деления отрезка пополам



Алгоритм:

1) вычислить середину отрезка: $c = \dfrac{a+b}{2}$
2) если на отрезке $[a,c]$ есть решение, присвоить **`b:=c`**, иначе **`a:=c`**
3) повторять шаги 1-2 до тех пор, пока $b-a > 2\varepsilon$

? Что напоминает?

? п.2: как определить, есть ли решение?

$$f(a)\cdot f(c) \le 0$$

Вариант: $\operatorname{sign}[f(a)] \ne \operatorname{sign}[f(c)]$

$$\operatorname{sign} x = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

Метод деления отрезка пополам

Алгоритмический язык:

```
delta:= 2*eps
нц пока b - a > delta
  c:= (a + b) / 2
  если f(a)*f(c) <= 0 то
      b:= c
    иначе
      a:= c
  все
кц
вывод 'x = ', (a+b)/2
```

```
sign(f(a)) <> sign(f(c))
```

? Как меняется длина отрезка?

? За сколько шагов уменьшится в 1000 раз?

Метод деления отрезка пополам

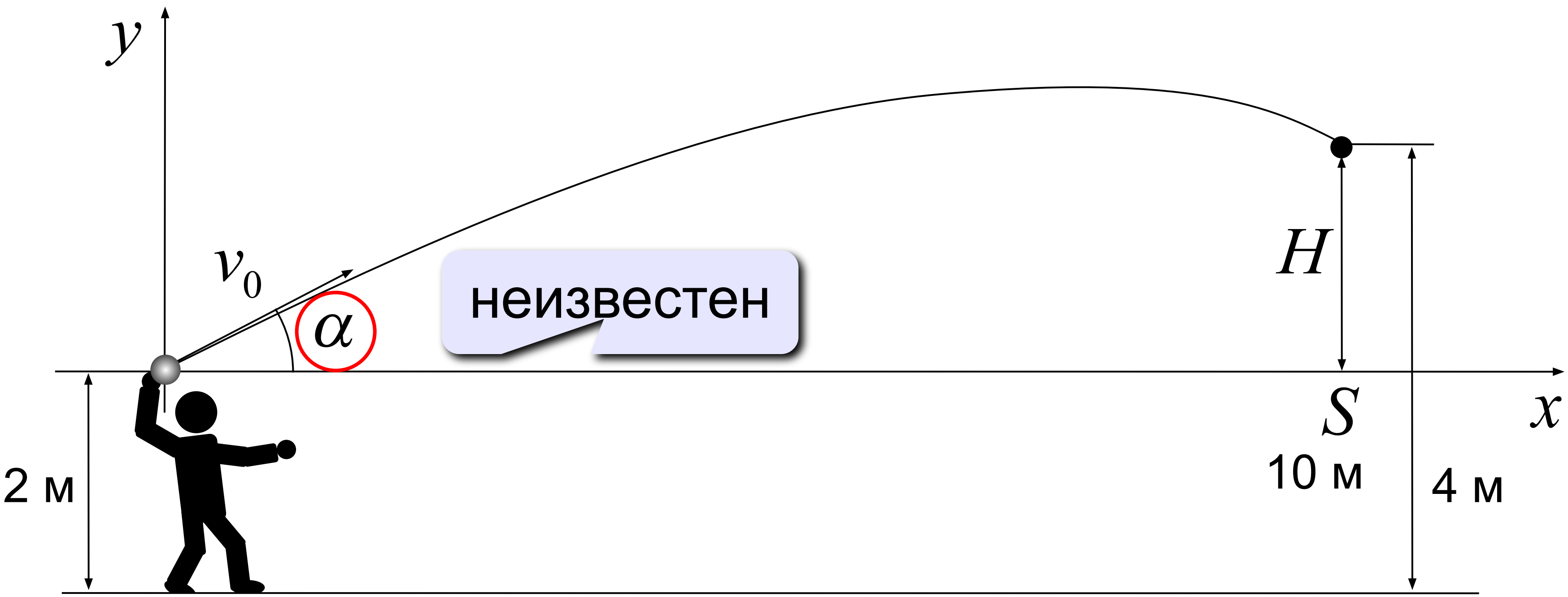
Паскаль:

```
delta:= 2*eps;
while b - a > delta do begin
  c:= (a + b) / 2;
  if f(a)*f(c) <= 0 then
       b:= c
  else a:= c;
end;
writeln('x = ', (a+b)/2:6:3);
```

? Как меняется длина отрезка?

? За сколько шагов уменьшится в 1000 раз?

Полёт мяча



$$x = v_0 \cdot t \cdot \cos\alpha,$$

$$y = v_0 \cdot t \cdot \sin\alpha - \frac{gt^2}{2}$$

Полёт мяча

Задача. Найти **угол α** (и время t) при котором $x = S$ и $y = H$:

$$S = v_0 \cdot t \cdot \cos\alpha, \quad H = v_0 \cdot t \cdot \sin\alpha - \frac{gt^2}{2}$$

Решение:

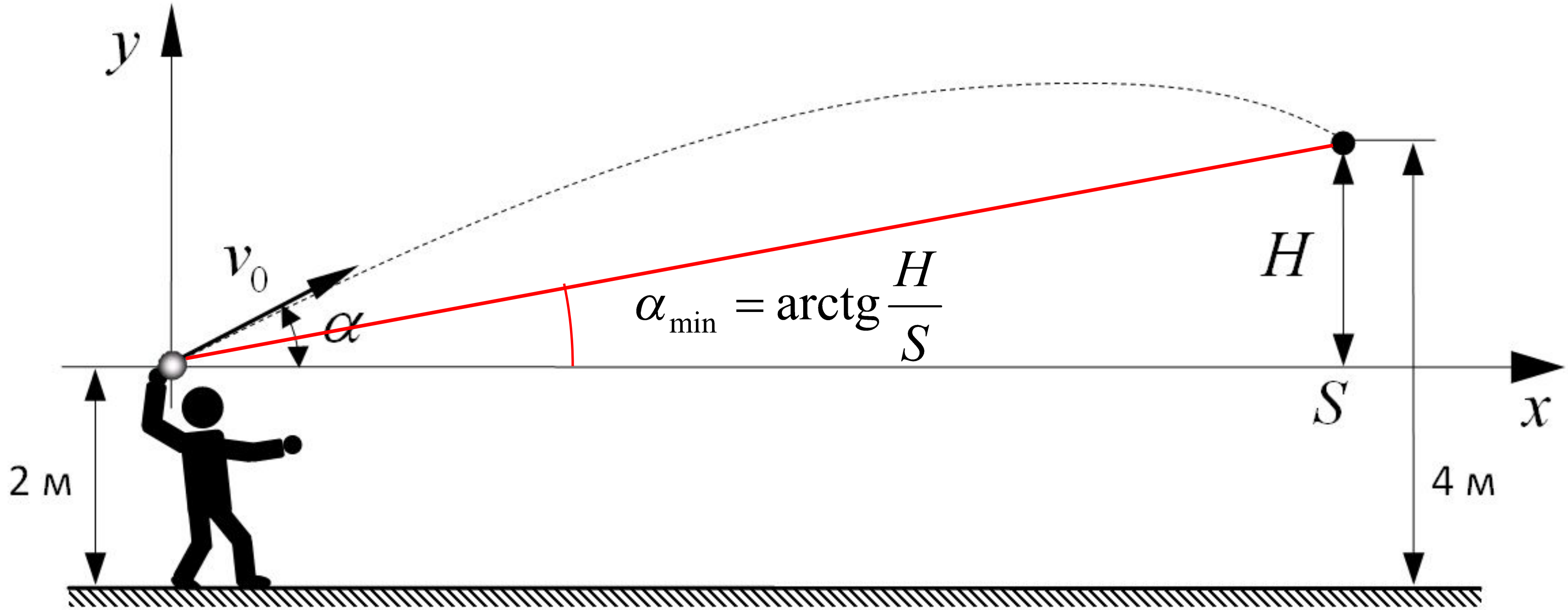
$$t = \frac{S}{v_0 \cdot \cos\alpha} \rightarrow H = \frac{\cancel{v_0} \cdot S \cdot \sin\alpha}{\cancel{v_0} \cdot \cos\alpha} - \frac{g \cdot S^2}{2v_0^2 \cos^2\alpha}$$

$$f(\alpha) = S \cdot \operatorname{tg}\alpha - \frac{g \cdot S^2}{2v_0^2 \cos^2\alpha} - H = 0$$

Диапазон углов для поиска: $[0^\circ ... 90^\circ] \Rightarrow \left[0 ... \frac{\pi}{2}\right]$

? Как уточнить?

Уточнение диапазона углов



Диапазон углов для поиска: $\left[\operatorname{arctg}\dfrac{H}{S}\ldots\dfrac{\pi}{2}\right]$

Полёт мяча

Программа на алгоритмическом языке:

```
pi:= 3.1415926
u:= 0
delta:= 2*eps
нц пока u < pi/2
  если f(u)*f(u+delta) <= 0 то
    вывод 'Угол: ', (u+eps)*180/pi
    вывод ' градусов', нс
  все
  u:= u + delta
кц
```

$$\alpha_1 \approx 35{,}6^\circ \qquad \alpha_2 \approx 65{,}8^\circ$$

Полёт мяча

Программа на языке Паскаль:

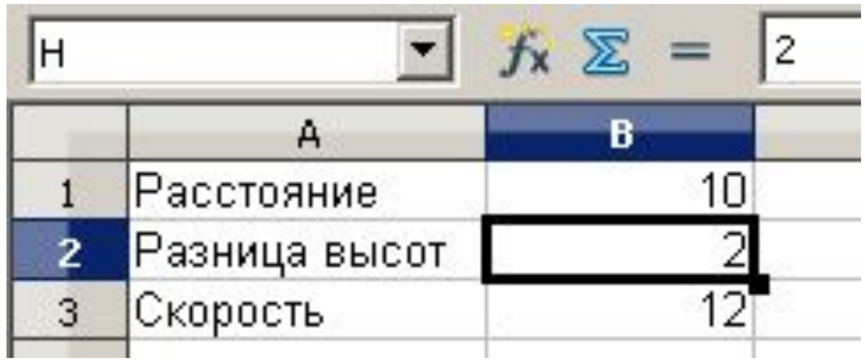
```
u:= 0;
delta:= 2*eps;
while u < pi/2 do begin
  if f(u)*f(u+delta) <= 0 then begin
    alpha:= (u+eps)*180/pi;
    writeln('Угол: ', alpha:4:1, ' градусов');
  end;
  u:= u + delta
end;
```

$$\alpha_1 \approx 35{,}6^\circ$$
$$\alpha_2 \approx 65{,}8^\circ$$

Полёт мяча

Использование табличного процессора:

— имя ячейки
или диапазона



Диапазон углов:

Полёт мяча

	A	B	C	D	E
1	Расстояние	10			
2	Разница высот	2			
3	Скорость	12			
4					
5	Угол	Угол(рад)	Время	y	f(alpha)
6	0	=RADIANS(A6)	=S/v/COS(B6)	=v*SIN(B6)*C6-9,81*C6^2/2	=D6-H
7	5				
8	10				

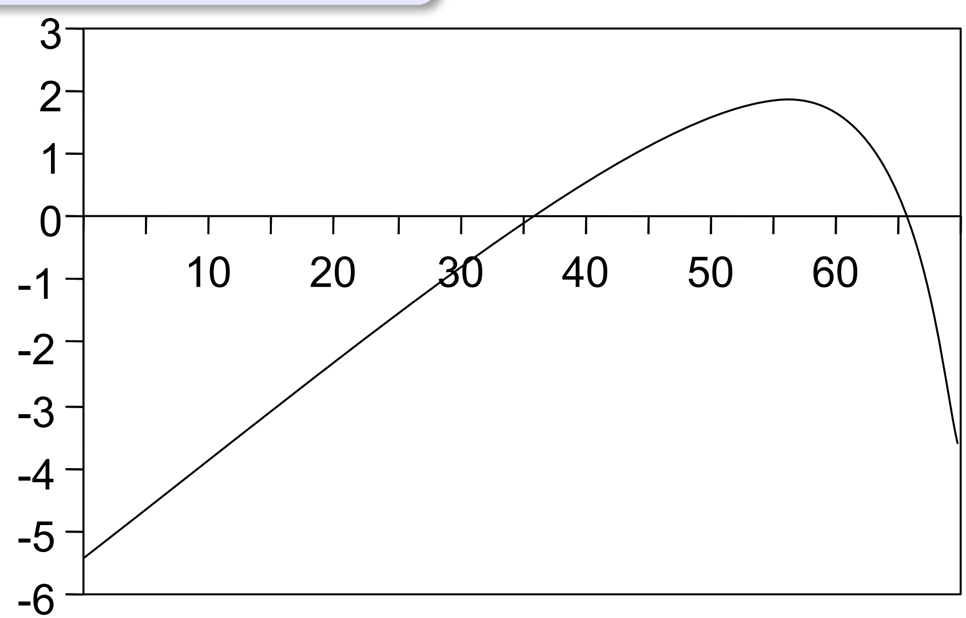
S ⇔ $B$1

Excel: **РАДИАНЫ**

Диаграмма XY:

Excel: **Точечная**

$$\alpha_1 \approx 35^\circ$$
$$\alpha_2 \approx 65^\circ$$



Полёт мяча

с графика!

	H	I	J	K	L
1	Угол	Угол(рад)	Время	y	f(alpha)
2	35				

начальное приближение

целевая ячейка

Сервис – Подбор параметра:

нужно $f(\alpha) = 0$

Как найти второе решение?

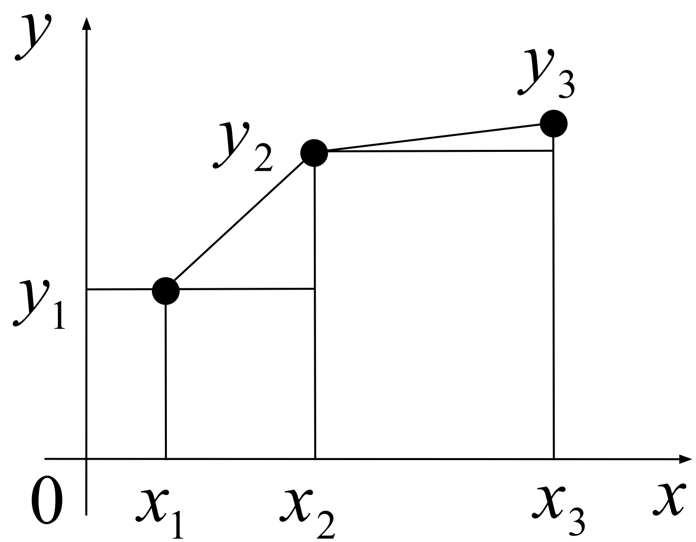
изменяем начальное приближение → результат в **H2**!

Решение вычислительных задач на компьютере

§ 71. Дискретизация

Вычисление длины линии

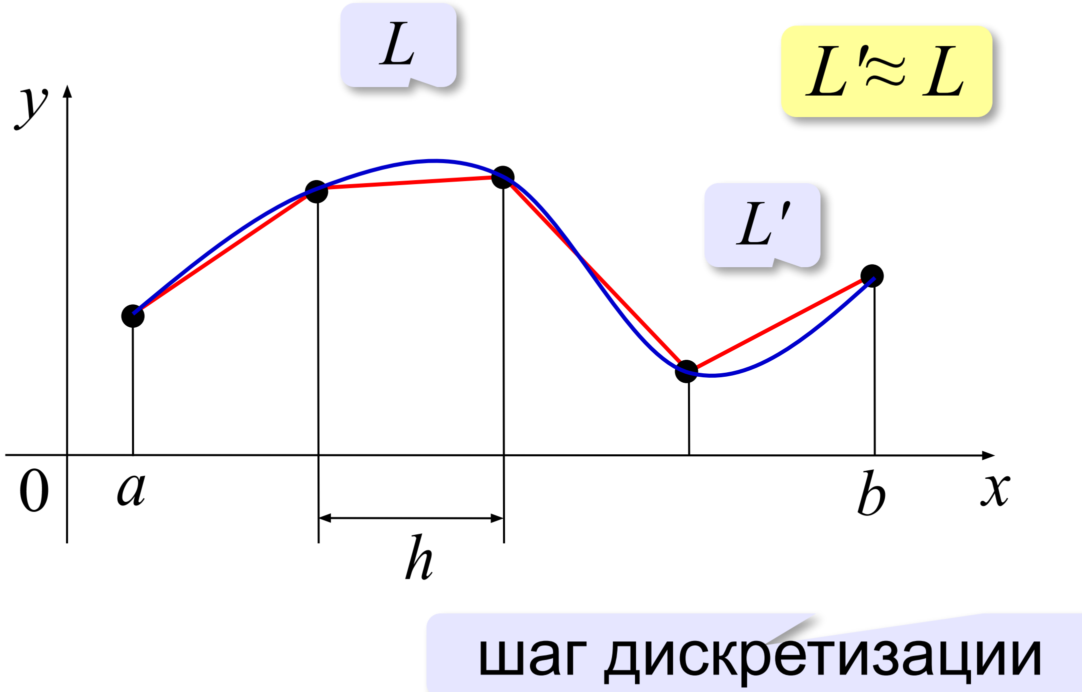
Ломаная:



$$L = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} + \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2}$$

Вычисление длины линии

Кривая:



! Выполнена дискретизация!

? Как увеличить точность? → $\downarrow h$

Дискретизация

- цель – представить задачу в виде, пригодном **для компьютерных расчётов**
- есть **потеря информации**
- методы **приближённые**
- для уменьшения погрешности нужно **уменьшать шаг дискретизации**

? Что ухудшится?

- при малом шаге на результат могут сильно влиять **погрешности вычислений**

Вычисление длины кривой

Программа на алгоритмическом языке:

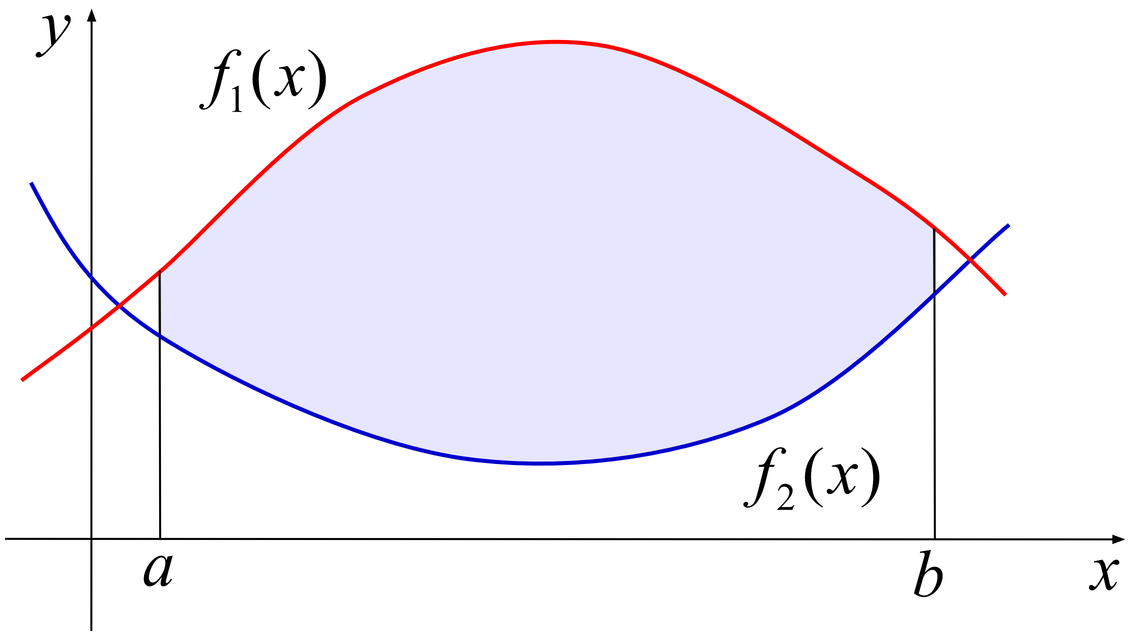
```
x:= a
L:= 0
нц пока x < b
  y1:= f(x)
  y2:= f(x+h)
  L:= L + sqrt(h*h + (y1-y2)*(y1-y2))
  x:= x + h
кц
вывод 'Длина кривой ', L
```

Вычисление длины кривой

Программа на Паскале:

```
x:= a;
L:= 0;
while x < b do begin
  y1:= f(x);
  y2:= f(x+h);
  L:= L + sqrt(h*h + (y2-y1)*(y2-y1));
  x:= x + h
end;
writeln('Длина кривой ', L:10:3);
```

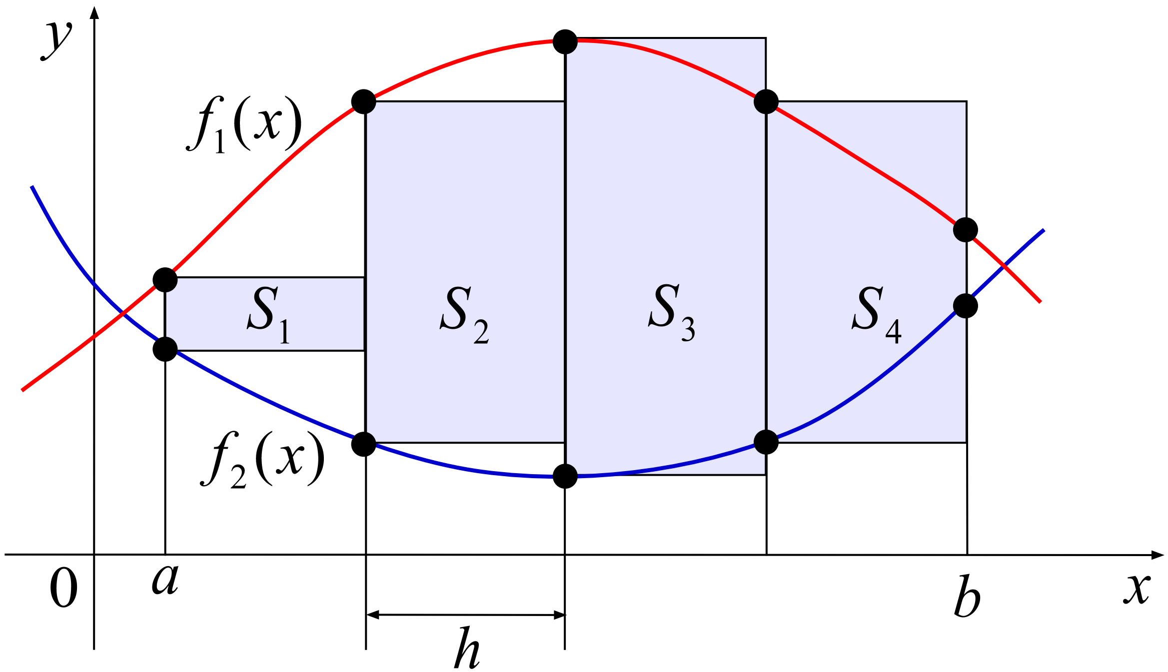
Площадь фигуры

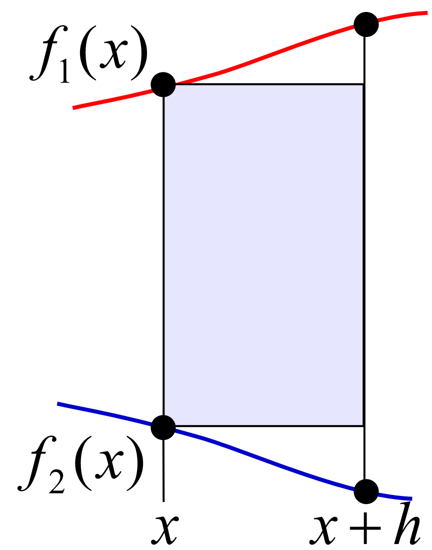


! Аналитически решается не всегда!

Дискретизация

Метод прямоугольников:





$$S \approx S_1 + S_2 + S_3 + S_4$$

$$S_x \approx h \cdot \left(f_1(x) - f_2(x)\right)$$

? Как улучшить?

Метод прямоугольников

Алгоритмический язык:

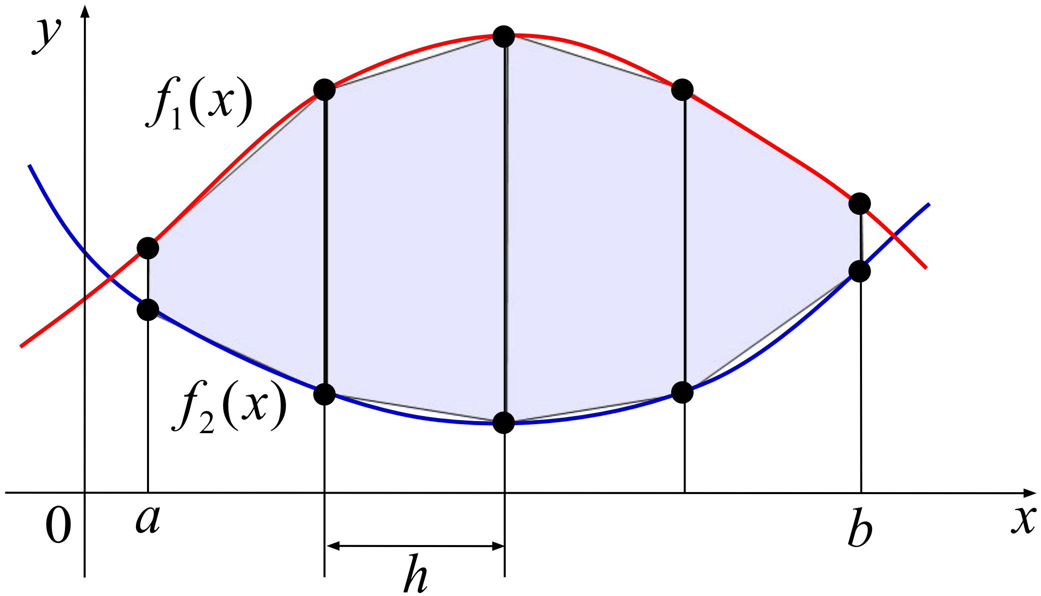
```
S:= 0; x:= a
нц пока x < b
  S:= S + f1(x+h/2) - f2(x+h/2)
  x:= x + h
кц
S:= h*S
вывод 'Площадь ', S
```

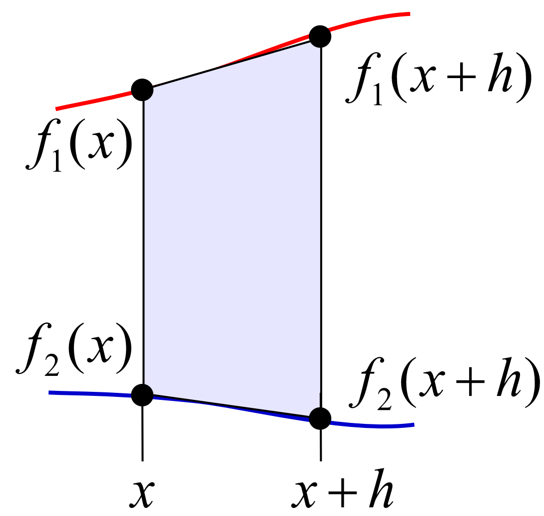
в середине отрезка $[x, x+h]$

Паскаль:

```
S:= 0; x:= a;
while x < b do begin
  S:= S + f1(x+h/2)- f2(x+h/2);
  x:= x + h
end;
S:= h*S;
writeln('Площадь ', S:8:3);
```

Метод трапеций





$$S_x = \frac{h}{2} \cdot \left[f_1(x) - f_2(x) + f_1(x+h) - f_2(x+h) \right]$$

Метод трапеций

Алгоритмический язык:

```
S:= 0; x:= a
нц пока x < b
  S:= S + f1(x) - f2(x) + f1(x+h) - f2(x+h)
  x:= x + h
кц
S:= h*S/2
вывод 'Площадь ', S
```

Паскаль:

? Как улучшить?

```
S:= 0; x:= a;
while x < b do begin
  S:= S + f1(x) - f2(x) + f1(x+h) -
  f2(x+h);
  x:= x + h;
end;
S:= h*S/2;
writeln('Площадь ', S:8:3);
```

Решение вычислительных задач на компьютере

§ 72. Оптимизация

Что такое оптимизация?

Оптимизация – это поиск наилучшего (*оптимального*) решения задачи в заданных условиях.

1) **Цель**: выбрать неизвестный x, так чтобы

$$f(x) \to \min \quad \text{или} \quad f(x) \to \max$$

$$-f(x) \to \min$$

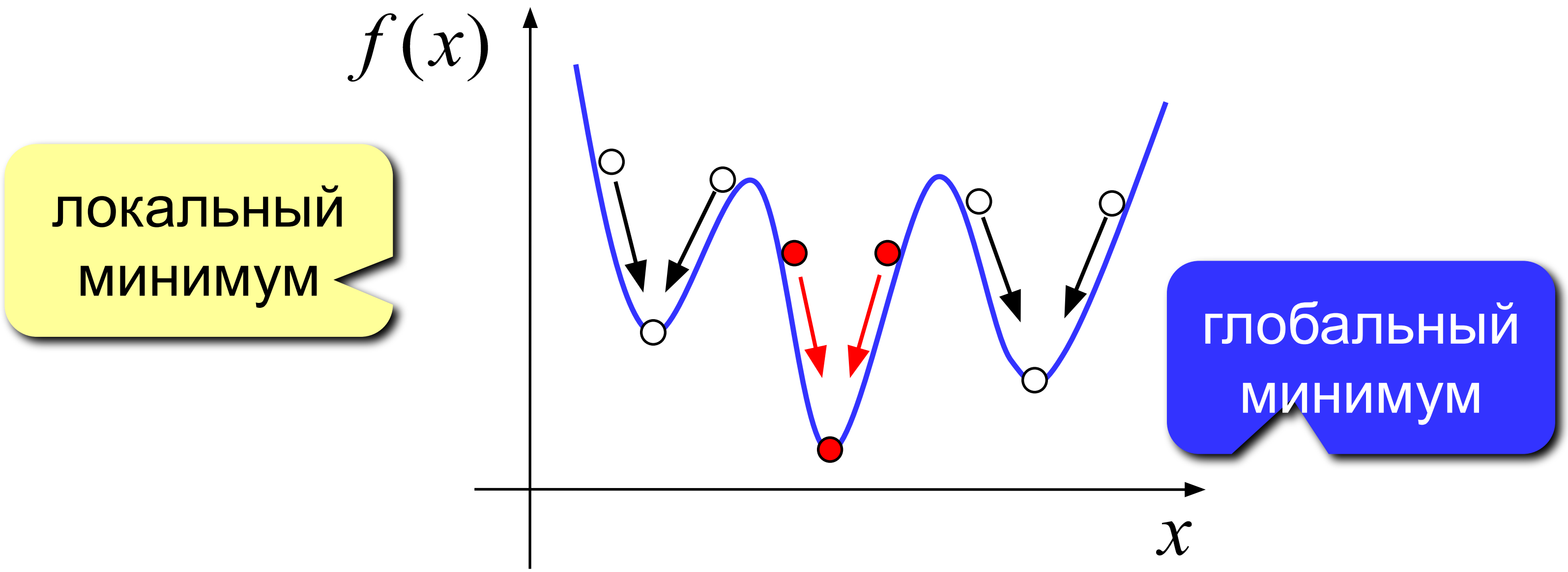
целевая функция

2) **Ограничения**

задача оптимизации

? Почему неправильно «самый оптимальный»?

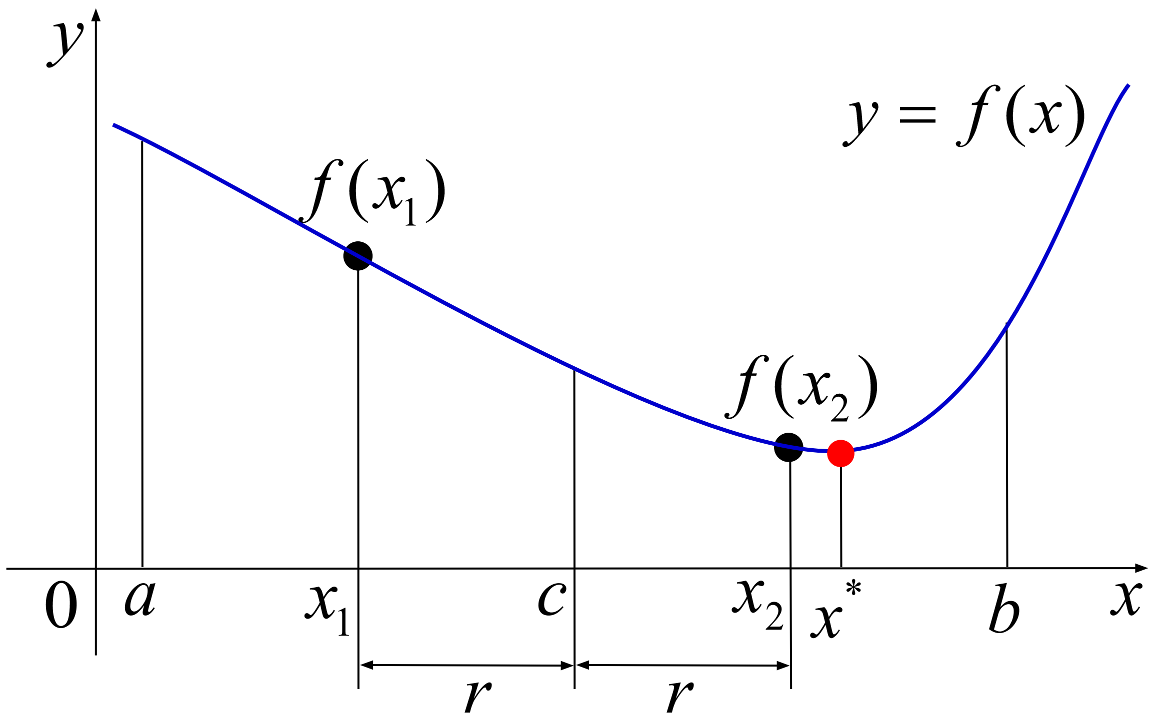
Что такое минимум?



- обычно нужно найти **глобальный** минимум
- большинство численных методов находят только **локальный** минимум

! Результат локальной оптимизации зависит от начального приближения!

Метод дихотомии



Алгоритм:

1) вычислить середину отрезка: $c = \dfrac{a+b}{2}$
2) найти симметричные точки $x_1 = c - r$, $x_2 = c + r$
3) если $f(x_1) > f(x_2)$, далее ищем на $[x_1, b]$

иначе ищем на $[a, x_2]$

Метод дихотомии

$$c = \frac{a+b}{2} \Rightarrow \quad x_1 = c - r, \quad x_2 = c + r$$

? Как выбрать r?

$$0 < r < \frac{b-a}{2} \Rightarrow \quad r = k \cdot (b-a), \quad 0 < k < 0{,}5$$

Уменьшение интервала:

стало

было

$$\frac{b-a}{2} + k \cdot (b-a) = (0{,}5 + k) \cdot \boxed{(b-a)}$$

! Выгоднее выбирать k близко к нулю!

? Почему нельзя $k = 0$?

Метод дихотомии

Алгоритмический язык:

```
k:= 0.01
delta:= 2*eps
нц пока b - a > delta
  r:= k*(b - a)
  x1:=(a + b)/2 - r
  x2:=(a + b)/2 + r
  если f(x1) > f(x2) то
        a:= x1
  иначе b:= x2
  все
кц
вывод 'x = ', (a+b)/2
```

? Как улучшить?

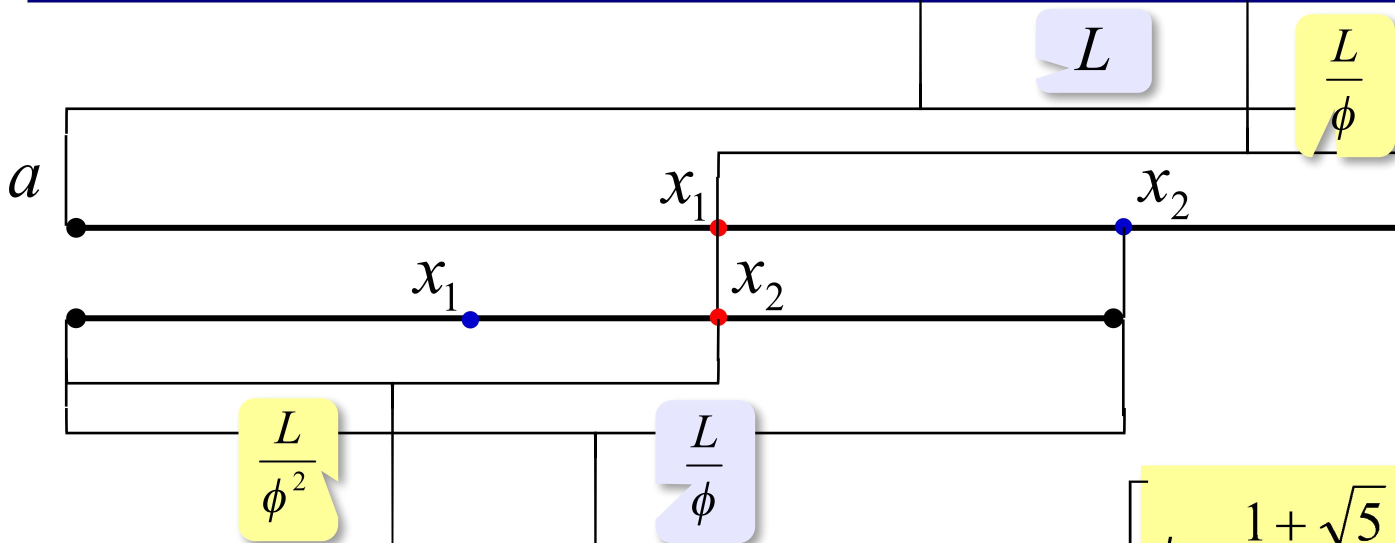
Метод дихотомии

Паскаль:

```
k:= 0.01;
delta:= 2*eps;
while b - a > delta do begin
  r:= k*(b - a);
  x1:=(a + b)/2 - r;
  x2:=(a + b)/2 + r;
  if f(x1) > f(x2) then
       a:= x1
  else b:= x2
end;
writeln('x = ', (a+b)/2:10:3 );
```

? Как улучшить?

Метод золотого сечения

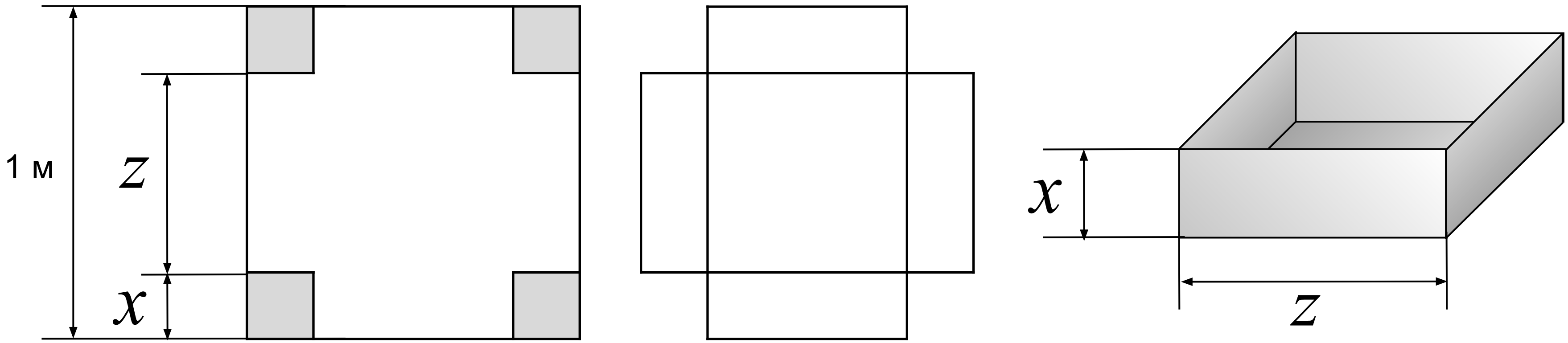


отношение **золотого сечения**

$$\frac{L}{\phi^2} = L - \frac{L}{\phi} \quad\Rightarrow\quad \phi^2 - \phi - 1 = 0 \quad\Rightarrow\quad \left[\begin{array}{l} \boxed{\phi = \dfrac{1+\sqrt{5}}{2} \approx 1{,}618\ldots} \\ \phi = \dfrac{1-\sqrt{5}}{2} \approx \cancel{-0{,}618\ldots} \end{array}\right.$$

$$\frac{(b-a)}{\phi} = (0{,}5+k)\cdot(b-a) \quad\Rightarrow\quad \boxed{0{,}5 + k = \frac{1}{\phi}}$$

Оптимальный раскрой листа



Цель: $V(x) \to \max$

$V(x) = x \cdot (1 - 2x)^2 \to \max$

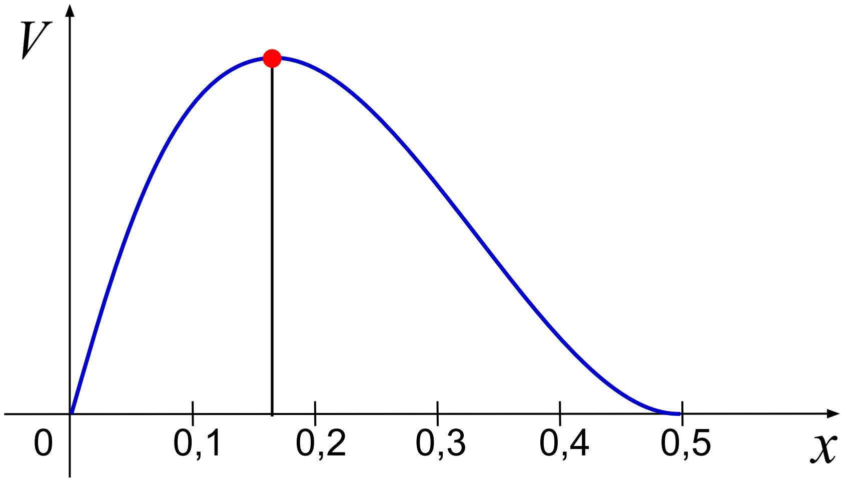
Какие ограничения?

Ограничения: $0 < x < 0{,}5$

Какой результат ожидаете (по интуиции)?

Оптимальный раскрой листа

В табличном процессоре:



начальное приближение ≈0,2

Какая формула в **F2**?

	E	F
1	x	Объем
2	0,200	0,072

Оптимизация в табличном процессоре

Задача оптимизации: найти максимум (или минимум) целевой функции в ячейке …, изменяя значения ячеек … при ограничениях ….

OpenOffice.org Calc:

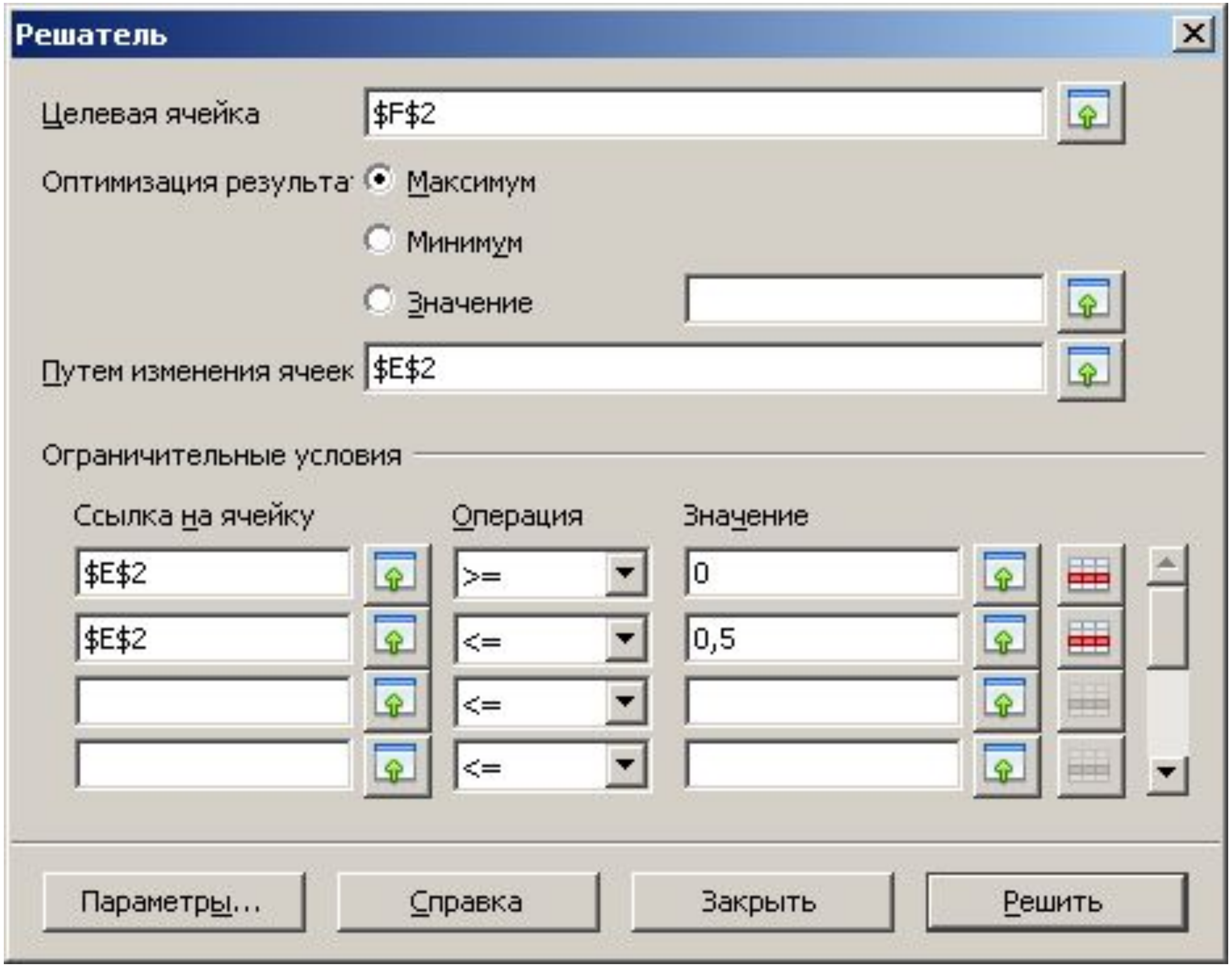
модуль ***Solver for Nonlinear Programming***
(входит в *LibreOffice*)

Excel:

надстройка ***Поиск решения***

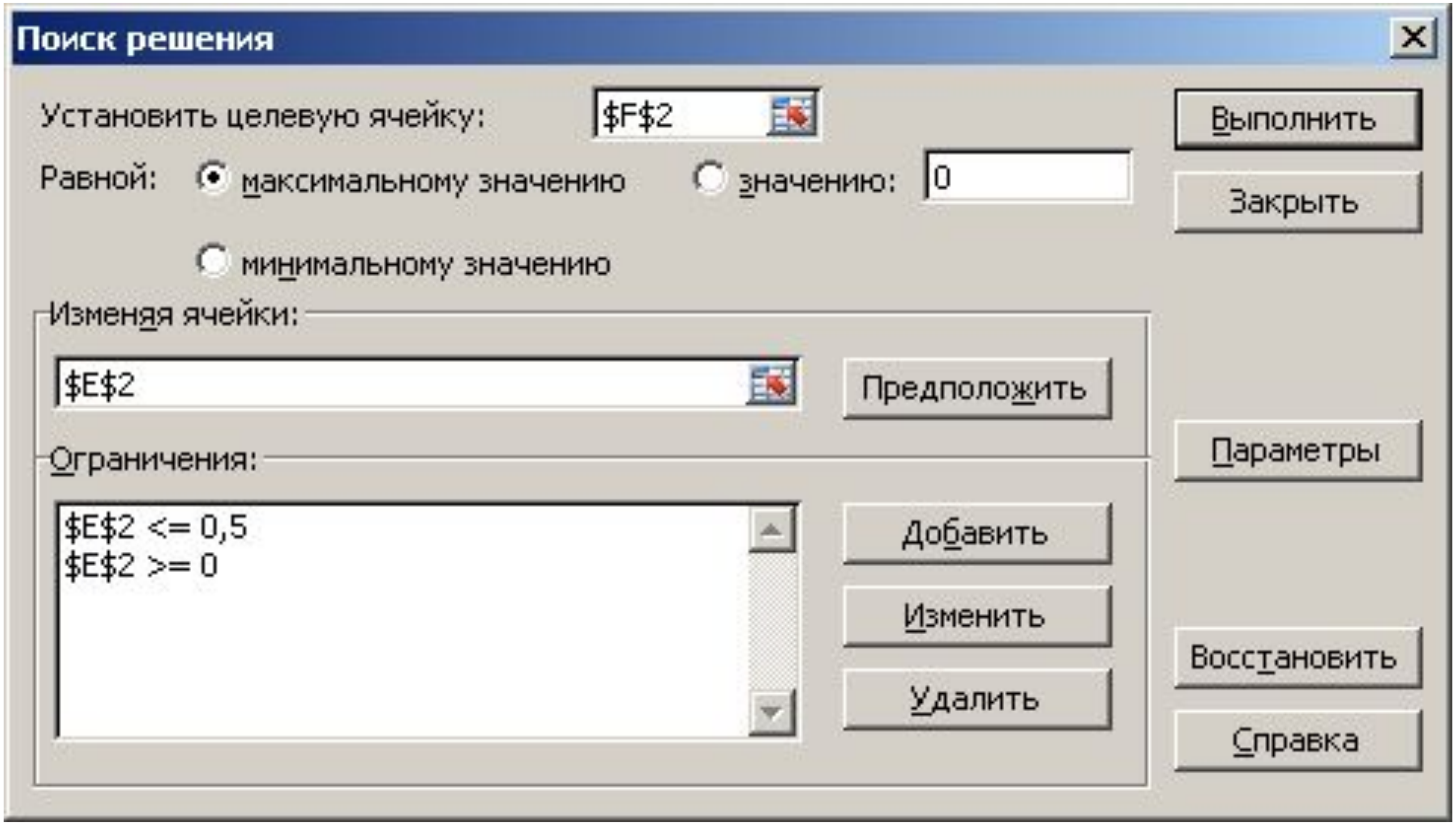
Оптимизация в табличном процессоре

OpenOffice.org Calc:



Оптимизация в табличном процессоре

Excel:



Решение вычислительных задач на компьютере

§ 73. Статистические расчёты

Что такое статистика?

Статистика – это наука, которая изучает методы обработки и анализа больших массивов данных.

Ряд данных: $x_1, x_2, \dots x_n$

только числа!

Свойства ряда данных:

сумма:	=SUM(A1:A20)	=СУММ(A1:A20)
среднее:	=AVERAGE(A1:A20)	=СРЗНАЧ(A1:A20)
минимальное:	=MIN(A1:A20)	=МИН(A1:A20)
максимальное:	=MAX(A1:A20)	=МАКС(A1:20)
количество чисел:	=COUNT(A1:A20)	=СЧЁТ(A1:20)

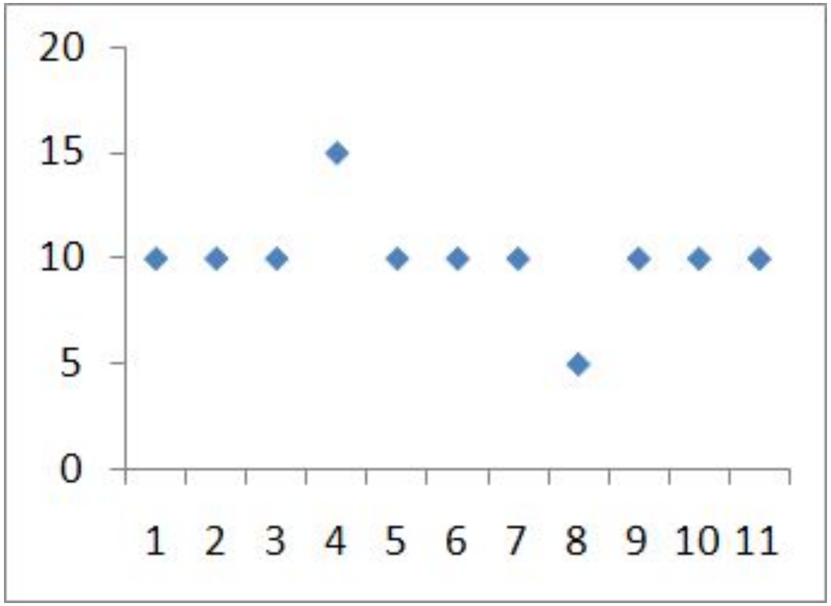
сколько ячеек удовлетворяет условию:

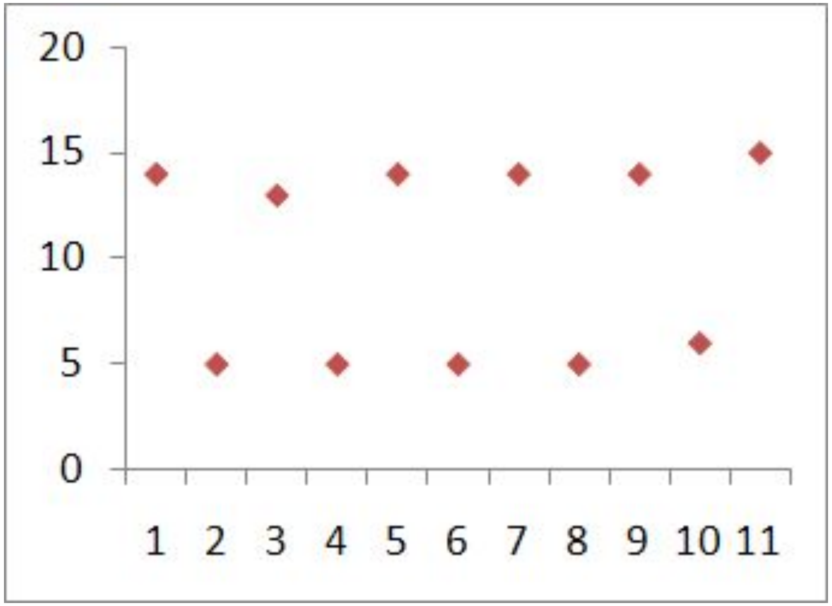
```
=COUNTIF(A1:A20;"=5")
=COUNTIF(A1:A20;">3")
```

СЧЁТЕСЛИ

Дисперсия

Для этих рядов одинаковы МИН, МАКС, СРЗНАЧ





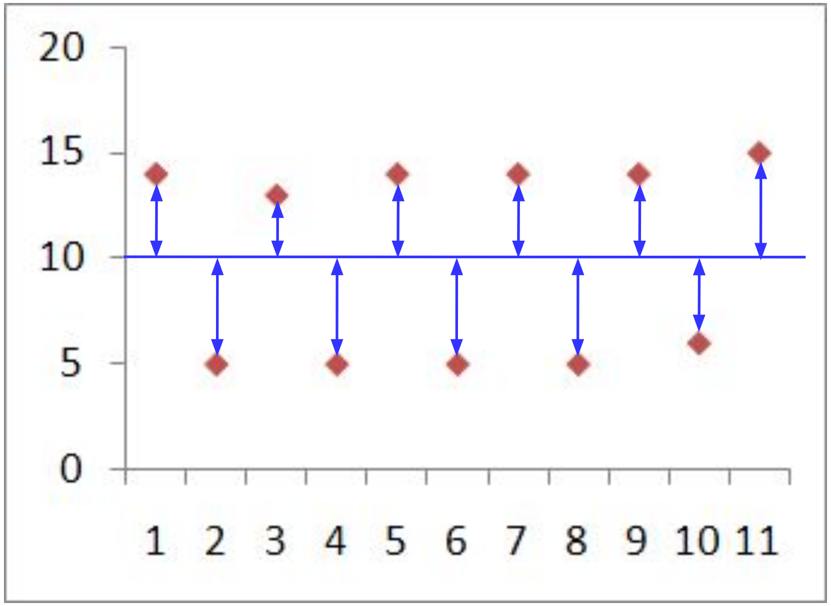
? В чем различие?

Дисперсия («разброс») характеризует разброс данных относительно среднего значения.

Дисперсия

$$D_x = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}$$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$ среднее арифметическое



$(x_1 - \bar{x})^2$ квадрат отклонения x_1 от среднего

D_x **средний квадрат отклонения** от среднего значения

Дисперсия и СКВО

$$D_x = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$$

? В каких единицах измеряется?

Что неудобно:

если x измеряется в метрах, то D_x – в м²

СКВО = среднеквадратическое отклонение

$$\sigma_x = \sqrt{D_x}$$

```
=STDEVP(A1:A20)
=СТАНДОТКЛОНП(A1:A20)
```

Условные вычисления

	A	B	C
1	Заказ	Сумма	Доставка
2	1234	256 руб.	51 руб.
3	1345	128 руб.	26 руб.
4	1456	1 024 руб.	0
5	1565	512 руб.	0
6	1576	345 руб.	69 руб.

Доставка:
- бесплатно при >500 руб.
- 20% для остальных

```
если B2 > 500 то
  C2:= 0
иначе
  C2:= B2*0.2
все
```

условие

```
=   IF(B2>500;0;B2*0,2)
=ЕСЛИ(B2>500;0;B2*0,2)
```

если «да»

если «нет»

Сложные условия

NOT (**НЕ**, отрицание)
AND (**И**, логическое умножение)
OR (**ИЛИ**, логическое сложение)

```
=IF(AND(A2<1500;B2>500);0;B2*0,2)
```

	A	B	C	D
1	**Фамилия**	**Год рождения**	**Рост**	**Принят**
2	Алексеев	1995	176	=IF(AND(B2>1994; C2>175);"да";"–")
3	Викторов	1995	167	=IF(AND(B3>1994; C3>175);"да";"–")
4	Петров	1994	180	=IF(AND(B4>1994; C4>175);"да";"–")

```
=IF(AND(B2>1994;C2>175); "да"; "–")
```

	A	B	C	D
1	**Фамилия**	**Год рождения**	**Рост**	**Принят**
2	Алексеев	1995	176	да
3	Викторов	1995	167	–
4	Петров	1994	180	–

Сложные условия

	A	B	C	D
1	Фамилия	Математика	Физика	Принят
2	Алексеев	100	67	=IF(OR(B2=100; C2=100;B2+C2>180);"да";"–")
3	Викторов	98	98	=IF(OR(B3=100; C3=100;B3+C3>180);"да";"–")
4	Петров	90	80	=IF(OR(B4=100; C4=100;B4+C4>180);"да";"–")

```
=IF(OR(B2=100;C2=100;B2+C2>=180); "да"; "–")
```

	A	B	C	D
1	Фамилия	Математика	Физика	Принят
2	Алексеев	100	67	да
3	Викторов	98	98	да
4	Петров	90	80	–

Вложенные условия

	A	B	C
1	Заказ	Сумма	Доставка
2	1234	256 руб.	26 руб.
3	1345	128 руб.	26 руб.
4	1456	1 024 руб.	0
5	1565	512 руб.	0
6	1576	345 руб.	35 руб.

Доставка:
- бесплатно при >500 руб.
- 10% при >200 руб.
- 20% для остальных

```
если B2 > 500 то
  C2:= 0
иначе
  если B2 > 200 то
    C2:=B2*0.1
  иначе
    C2:=B2*0.2
  все
все
```

```
=IF(B2>500;0;IF(B2>200;B2*0,1;B2*0,2))
```

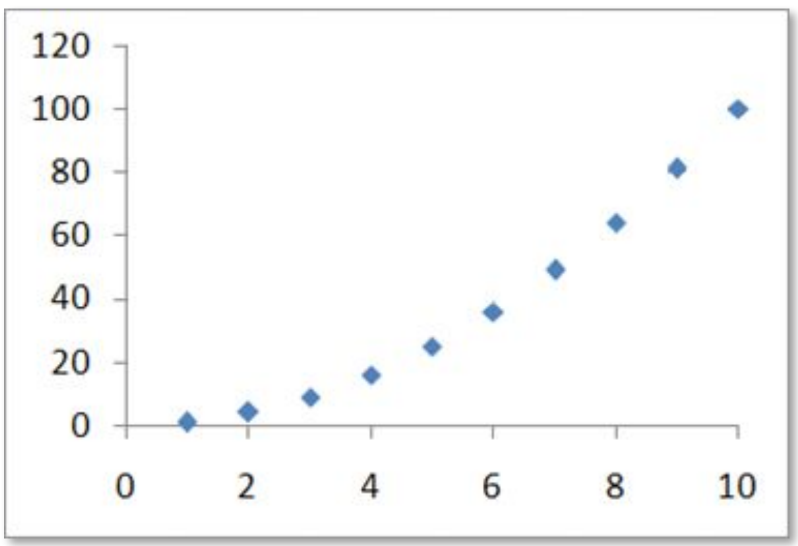
Связь двух рядов данных

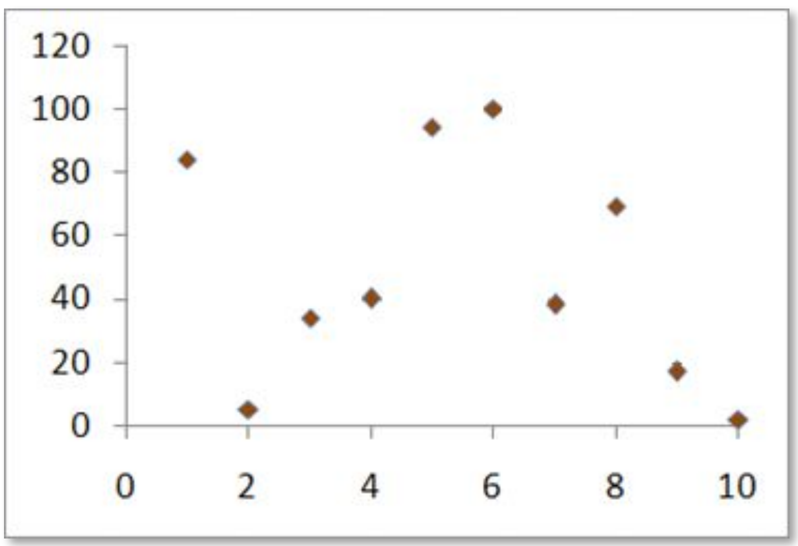
Два ряда одинаковой длины:

$$x_1,\ x_2,\ \ldots,\ x_n \qquad\qquad y_1,\ y_2,\ \ldots,\ y_n$$

Вопросы:

- есть ли связь между этими рядами (соответствуют ли пары (x_i, y_i) какой-нибудь зависимости $y = f(x)$)
- насколько сильна эта связь?





Коэффициент корреляции

средние рядов

$$\rho_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sigma_x \cdot \sigma_y}$$

СКВО рядов

? В каких единицах измеряется?

безразмерный!

? Если x и y – один и тот же ряд?

$$-1 \le \rho_{xy} \le 1$$

```
=CORREL(A1:A20;B1:B20)
=КОРРЕЛ(A1:A20;B1:B20)
```

Коэффициент корреляции

Как понимать это число?

- если $\rho_{xy} > 0$: увеличение x приводит к увеличению y
- если $\rho_{xy} < 0$: увеличение x приводит к уменьшению y
- если $\rho_{xy} \approx 0$: связь обнаружить не удалось

Сильная связь: $|\rho_{xy}| > 0{,}5$

$\rho_{xy} = 1$: линейная зависимость $y = kx + b, \quad k > 0$

$\rho_{xy} = -1$: линейная зависимость $y = kx + b, \quad k < 0$

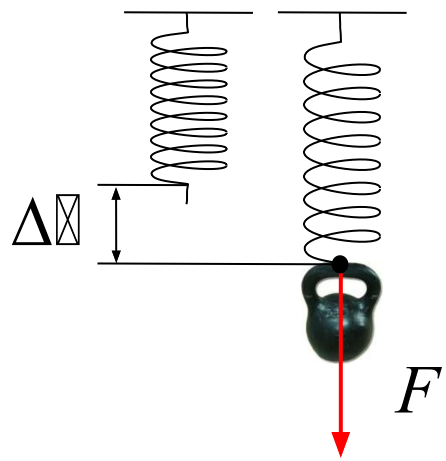
? Если $\rho_{xy} \approx 0$, то связи нет?

! Метод для определения линейной зависимости!

Решение вычислительных задач на компьютере

§ 74. Обработка результатов эксперимента

Закон Гука

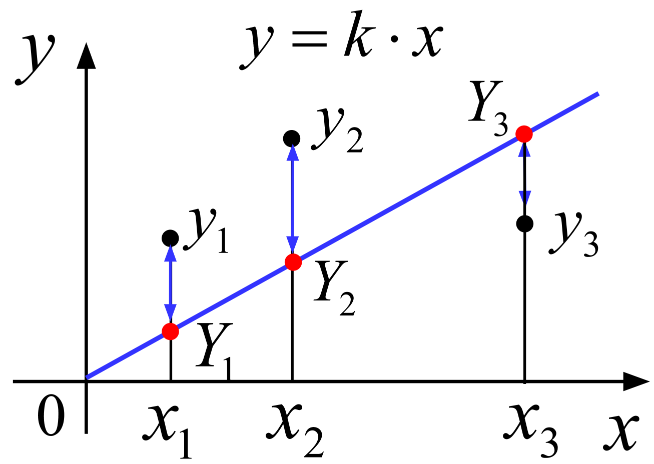


$$F = k \cdot \Delta l \quad \Rightarrow \quad k = \frac{F}{\Delta l}$$

Несколько опытов: $k_i = \dfrac{F_i}{\Delta l_i}, \quad (i = 1,\ldots n)$

? Что принять за k?

Метод наименьших квадратов (МНК)



неизвестно!

$$Y_i = k \cdot x_i$$

Ошибка определяется величиной:

$$E(k) = (Y_1 - y_1)^2 + (Y_2 - y_2)^2 + \ldots + (Y_n - y_n)^2 = \sum_{i=1}^{n} (Y_i - y_i)^2$$

Метод наименьших квадратов: $E(k) \to \min$

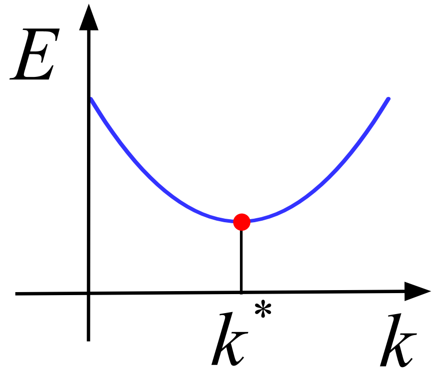
! Это задача оптимизации!

Метод наименьших квадратов (МНК)

$$(Y_i - y_i)^2 = (k \cdot x_i - y_i)^2 = k^2 x_i^2 - 2k x_i y_i + y_i^2$$

$$E(k) = A \cdot k^2 - B \cdot k + C$$

$$A = \sum_{i=1}^{n} x_i^2 \qquad B = 2\sum_{i=1}^{n} x_i y_i \qquad C = \sum_{i=1}^{n} y_i^2$$



$$k^* = \frac{B}{2A}$$

Метод наименьших квадратов (МНК)

Алгоритмический язык:

```
A:= 0; B:= 0
нц для i от 1 до N
  A:=A+x[i]*x[i]
  B:=B+x[i]*y[i]
кц
k:=B/A
```

```
вещ A, B, k
вещ x[1:N], y[1..N]
```

$$k^* = \frac{\sum_{i=1}^{N} x_i \cdot y_i}{\sum_{i=1}^{N} x_i^2}$$

Паскаль:

```
A:= 0; B:= 0;
for i:=1 to N do begin
  A:=A+x[i]*x[i];
  B:=B+x[i]*y[i]
end;
k:=B/A;
```

```
var A, B, k: real;
    x,y: array[1..N] of real;
```

Метод наименьших квадратов (МНК)

Табличный процессор:

начальное приближение

	A	B	C
1	k	1,000	
2	E	=SUMXMY2(B5:B7;C5:C7)	
3			
4	x	y	Y
5	1	1,1	=$B$1*A5
6	2	1,8	=$B$1*A6
7	3	3,5	=$B$1*A7

СУММКВРАЗН

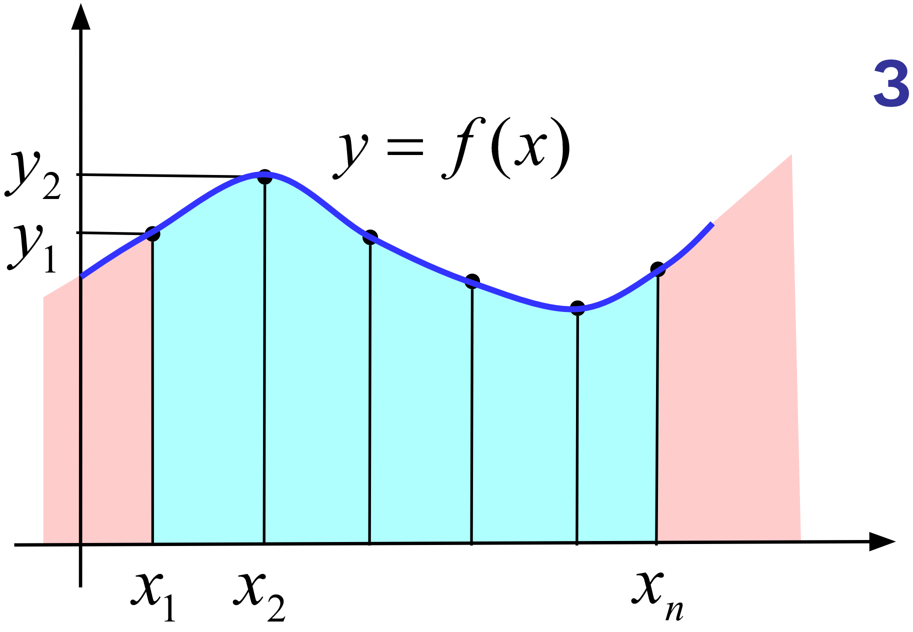
Поиск решения: выбрать **`B1`** так, что **`B2`**→**`min`**

Восстановление зависимостей

Два ряда одинаковой длины:

$$x_1,\ x_2,\ \ldots,\ x_n \qquad\qquad y_1,\ y_2,\ \ldots,\ y_n$$

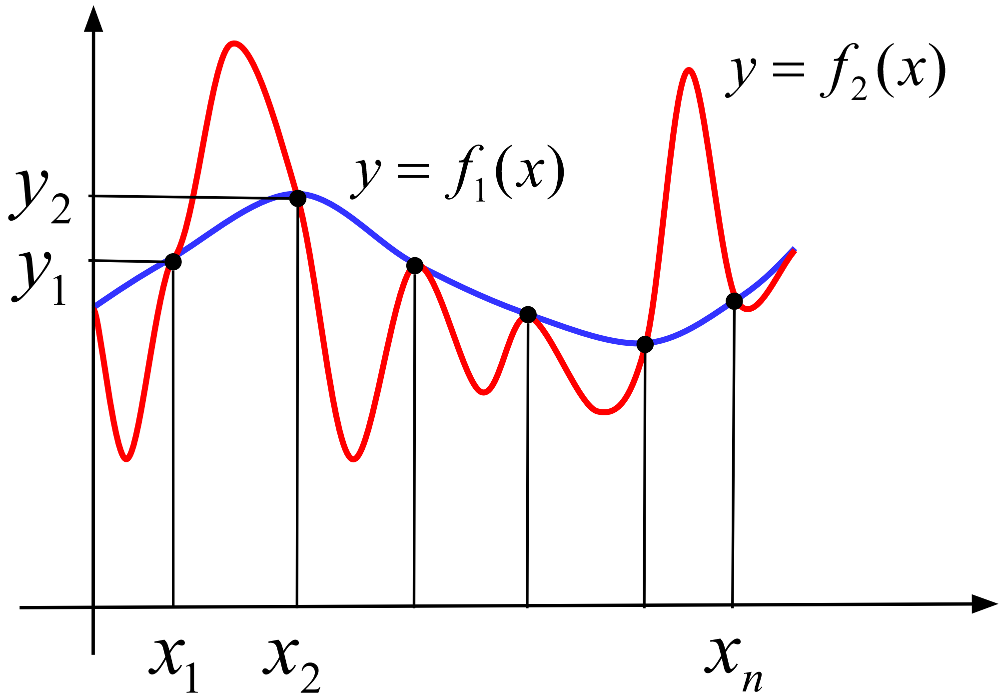
задают некоторую неизвестную функцию $y = f(x)$



Зачем:

- найти y в промежуточных точках (***интерполяция***)
- найти y вне диапазона измерений (***экстраполяция, прогнозирование***)

Восстановление зависимостей

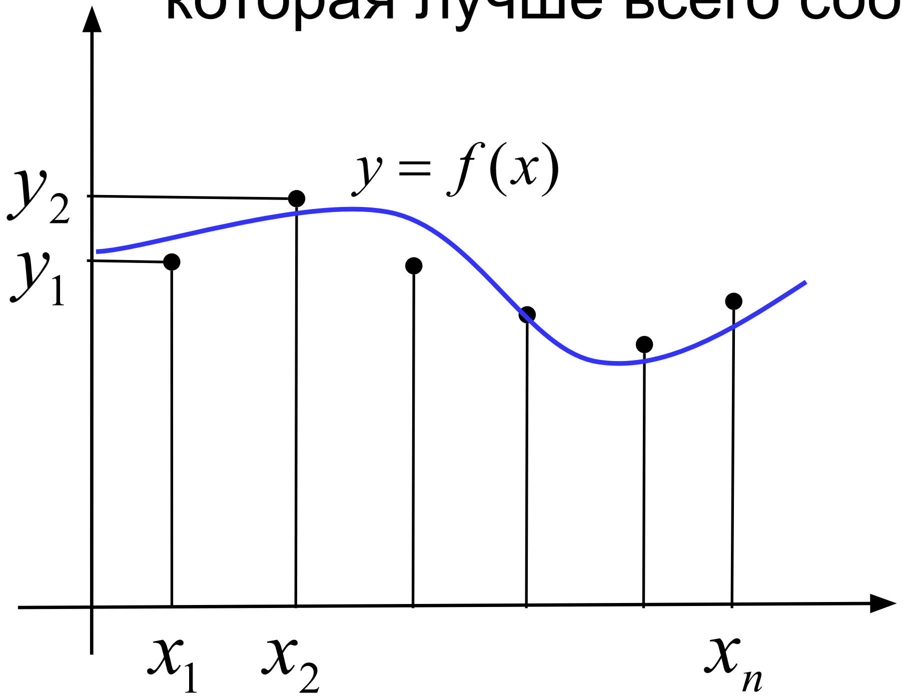


Через заданный набор точек проходит бесконечно много разных кривых!

Вывод: задача **некорректна**, поскольку решение неединственно.

Восстановление зависимостей

Корректная задача: найти функцию заданного вида, которая лучше всего соответствует данным.



Примеры:

- линейная $y = a \cdot x + b$
- полиномиальная $y = a_3 x^3 + a_2 x^2 + a_1 x + a_0$
- степенная $y = a \cdot x^b$
- экспоненциальная $y = a \cdot e^{bx}$
- логарифмическая $y = a \cdot \ln x + b$

! График функции не обязательно проходит через заданные точки!

? Как выбрать функцию?

Что значит «лучше всего соответствует»?

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - Y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

(x_i, y_i) заданные пары значений

$Y_i = f(x_i)$

$\overline{y}$ – среднее значение y_i

коэффициент детерминации

Крайние случаи:

- если график проходит через точки: $R^2 = 1$
- если считаем, что y не меняется и $Y_i = \overline{y}$: $R^2 = 0$

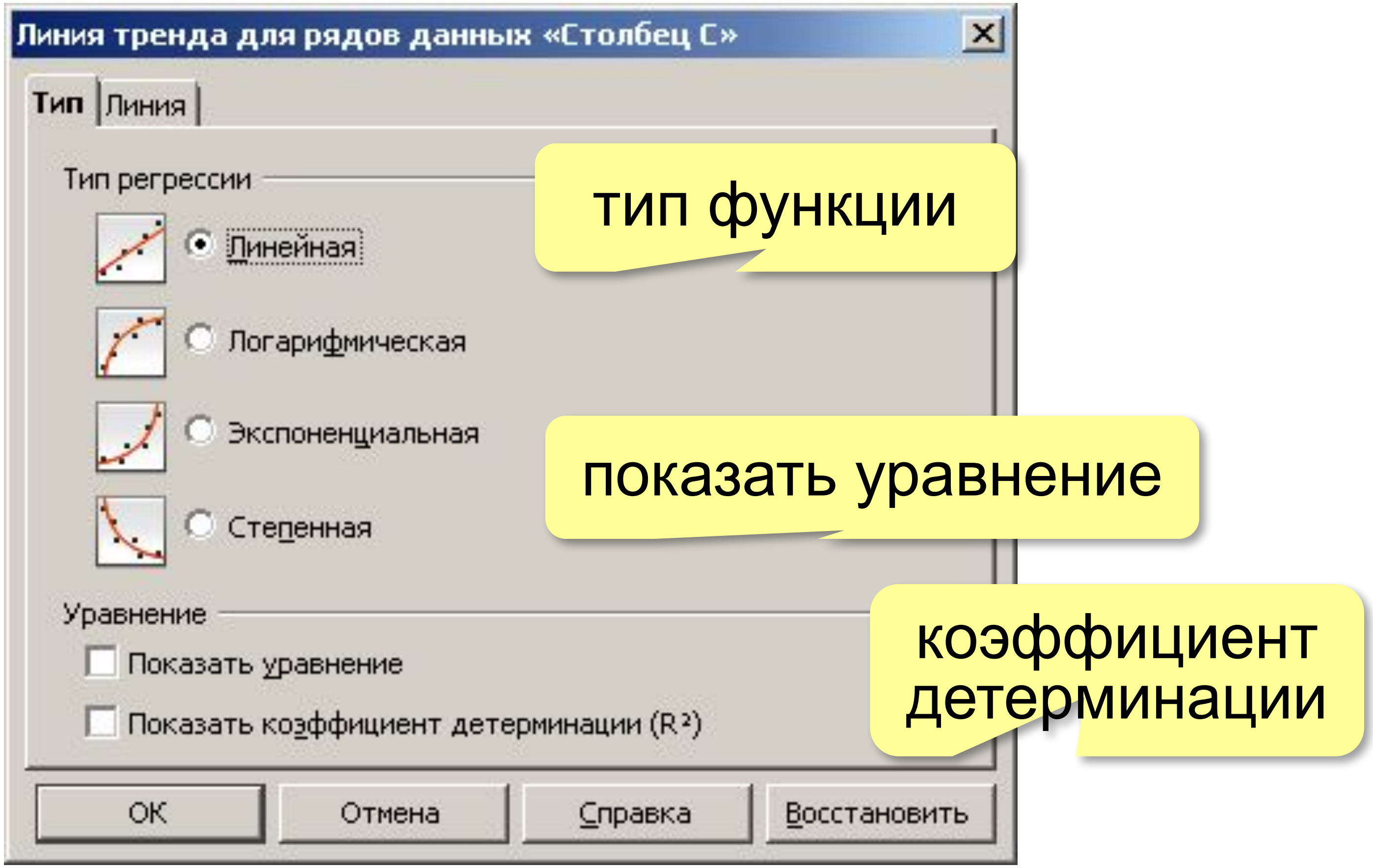
$$R^2 \to \max \text{ когда } \sum_{i=1}^{n}(y_i - Y_i)^2 \to \min$$

! Фактически – метод наименьших квадратов!

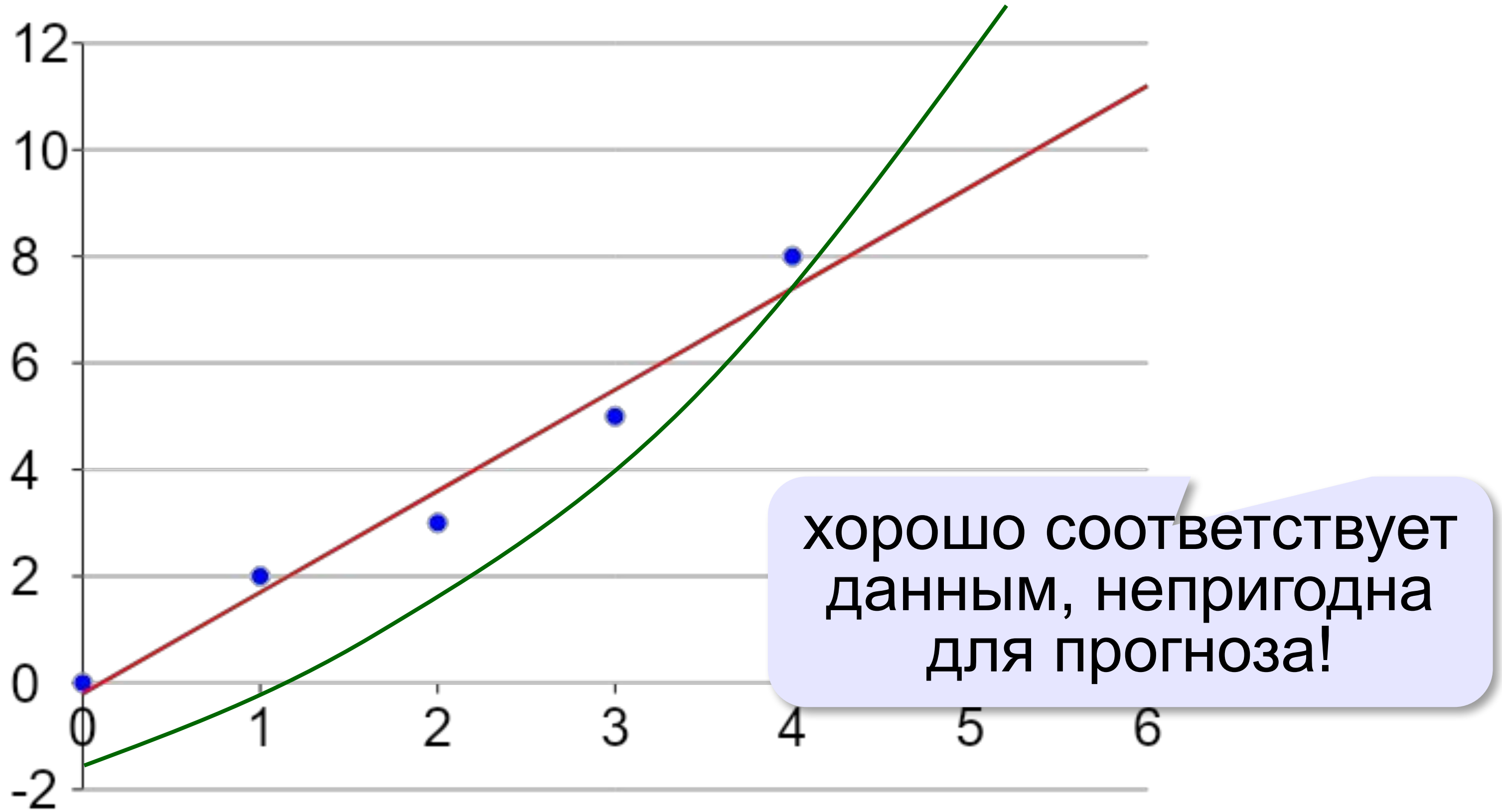
Восстановление зависимостей

Табличный процессор:

1) **Диаграмма XY** (*Excel:* Точечная)
2) ПКМ – Вставить линию тренда



Прогнозирование



Конец фильма

ПОЛЯКОВ Константин Юрьевич
д.т.н., учитель информатики
ГБОУ СОШ № 163, г. Санкт-Петербург
kpolyakov@mail.ru

ЕРЕМИН Евгений Александрович
к.ф.-м.н., доцент кафедры мультимедийной
дидактики и ИТО ПГГПУ, г. Пермь
eremin@pspu.ac.ru

Источники иллюстраций

1. vispo.ru
2. www.ars-sport.ru
3. иллюстрации художников издательства «Бином»
4. авторские материалы