Section 7.5

The Method of Least Squares

Best fit line

- Suppose that two quantities x and y are related approximately linearly, $y = Ax + B$, and we have some data →

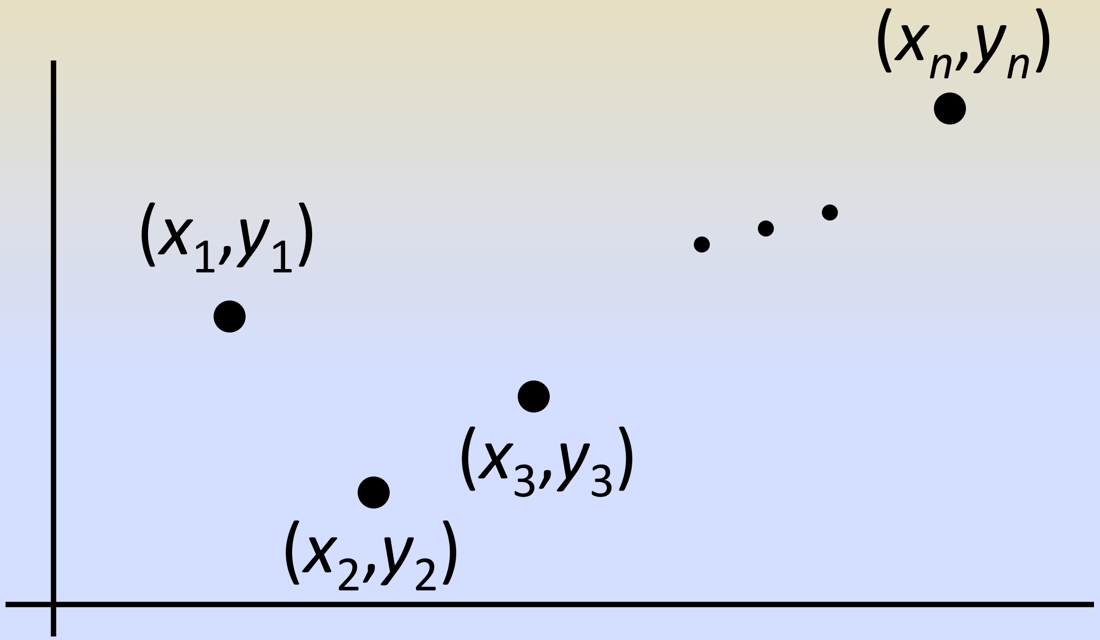
x	y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

Best fit line

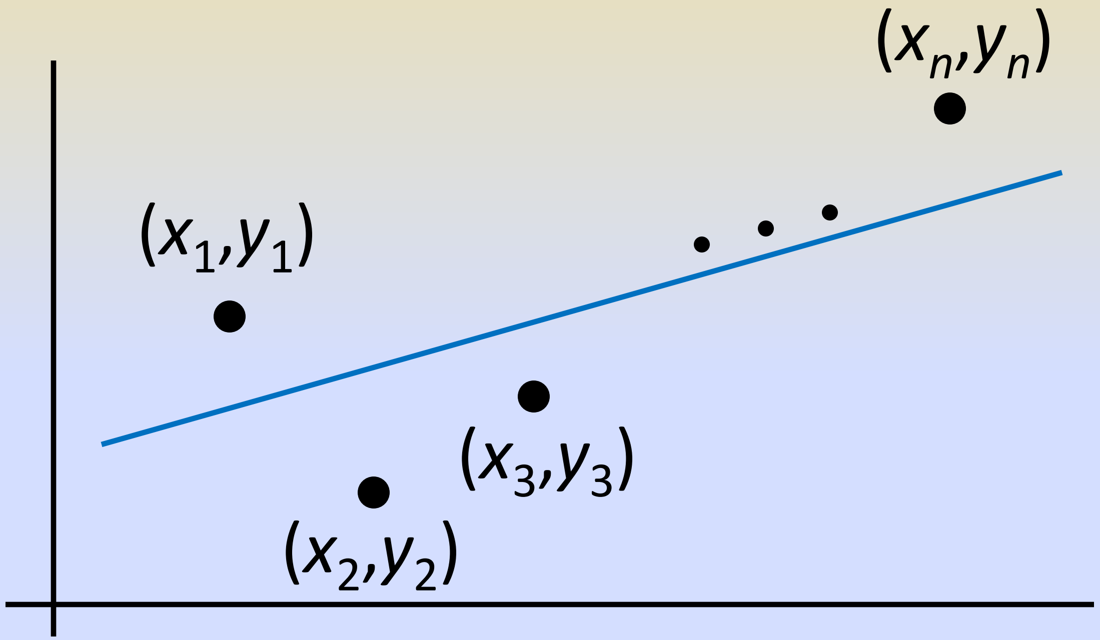
- Suppose that two quantities x and y are related approximately linearly, $y = Ax + B$, and we have some data →
- We want to **find the line** (that is, the slope A and the y-intercept B) **that "best" fits that data**.

x	y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

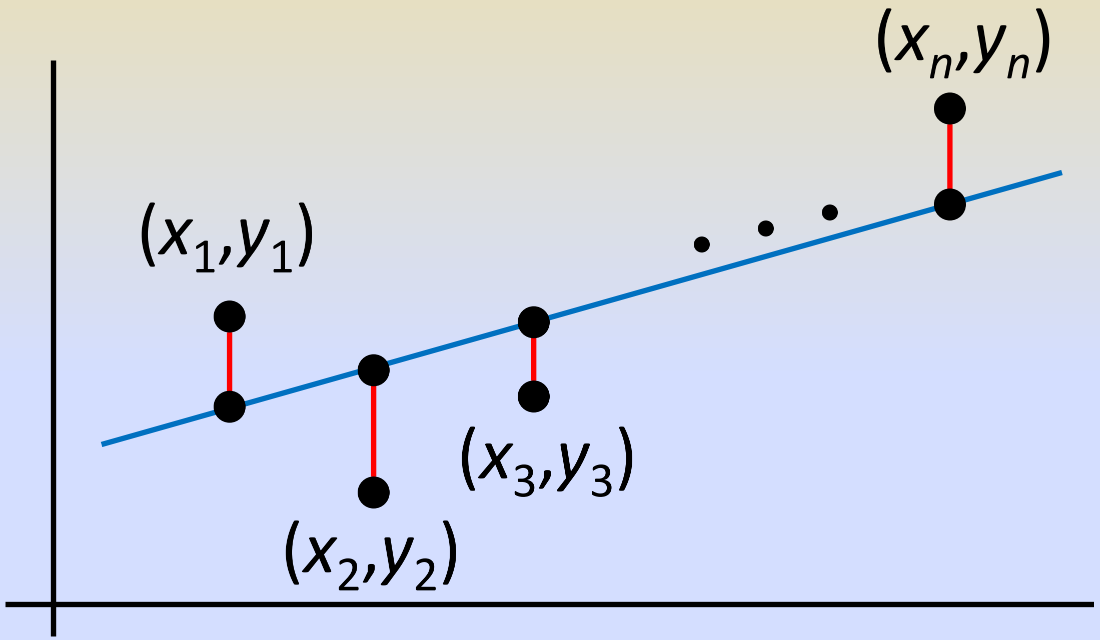
Best fit line $y = Ax + B$



Best fit line $y = Ax + B$

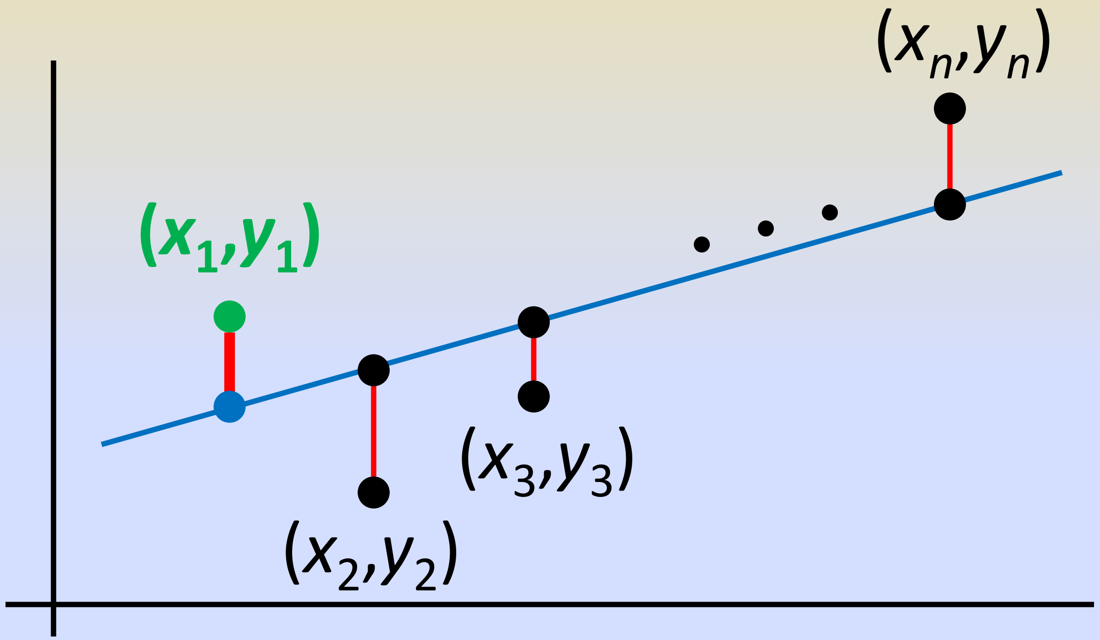


Best fit line $y = Ax + B$



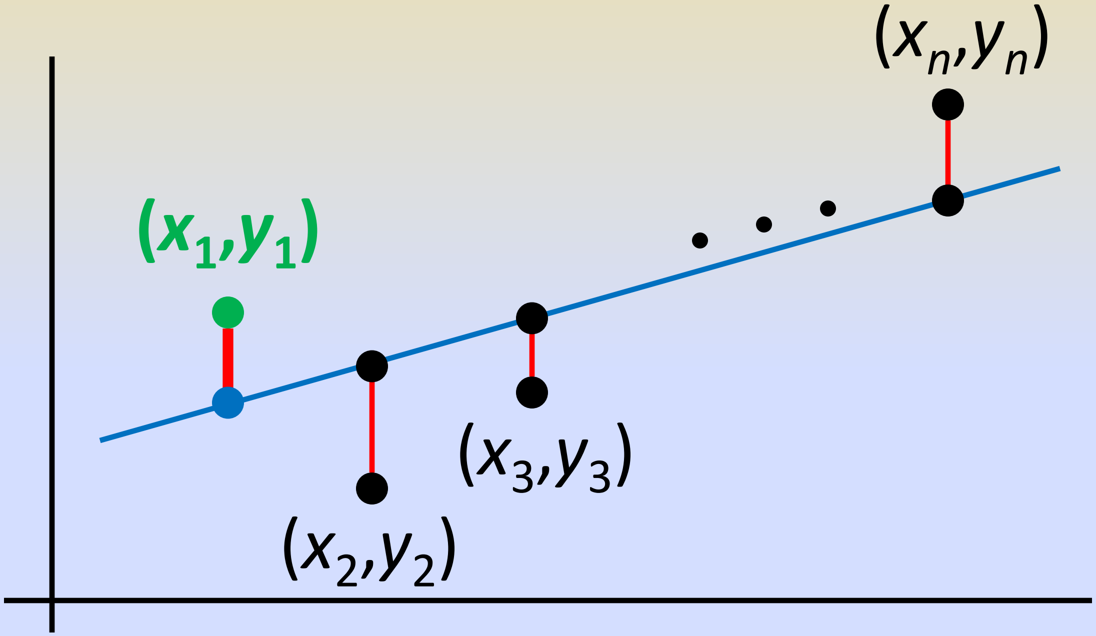
$$E = |Ax_1+B - y_1| + |Ax_2+B - y_2| + \dots + |Ax_n+B - y_n|$$

Best fit line $y = Ax + B$



$$E = |Ax_1 + B - y_1| + |Ax_2 + B - y_2| + \ldots + |Ax_n + B - y_n|$$

Best fit line $y = Ax + B$

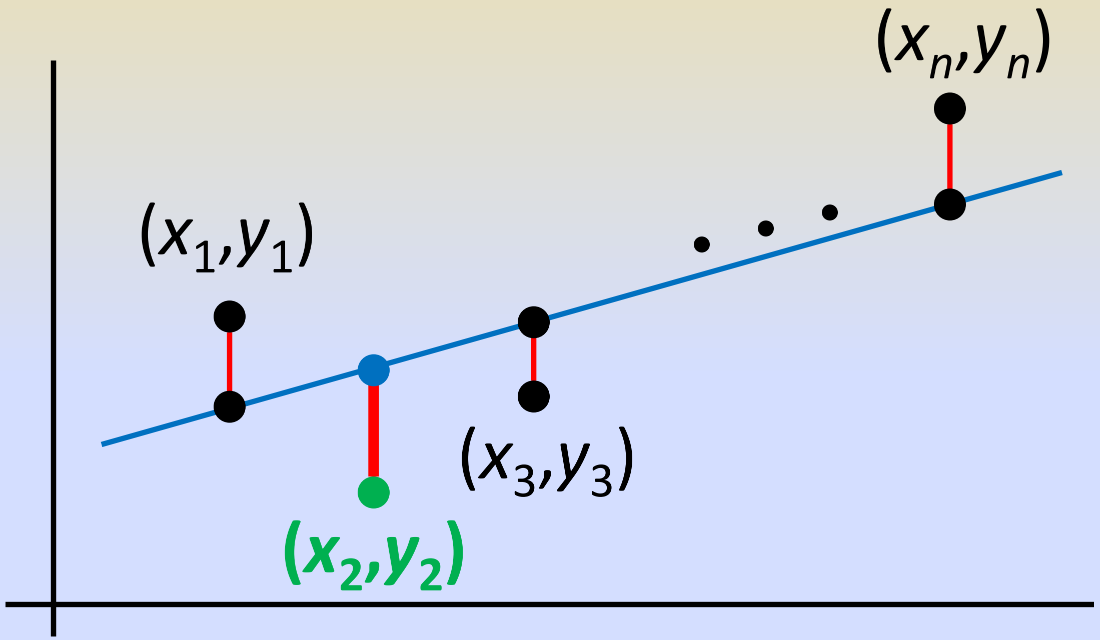


$$E = |Ax_1+B - y_1| + |Ax_2+B - y_2| + \ldots + |Ax_n+B - y_n|$$

↑

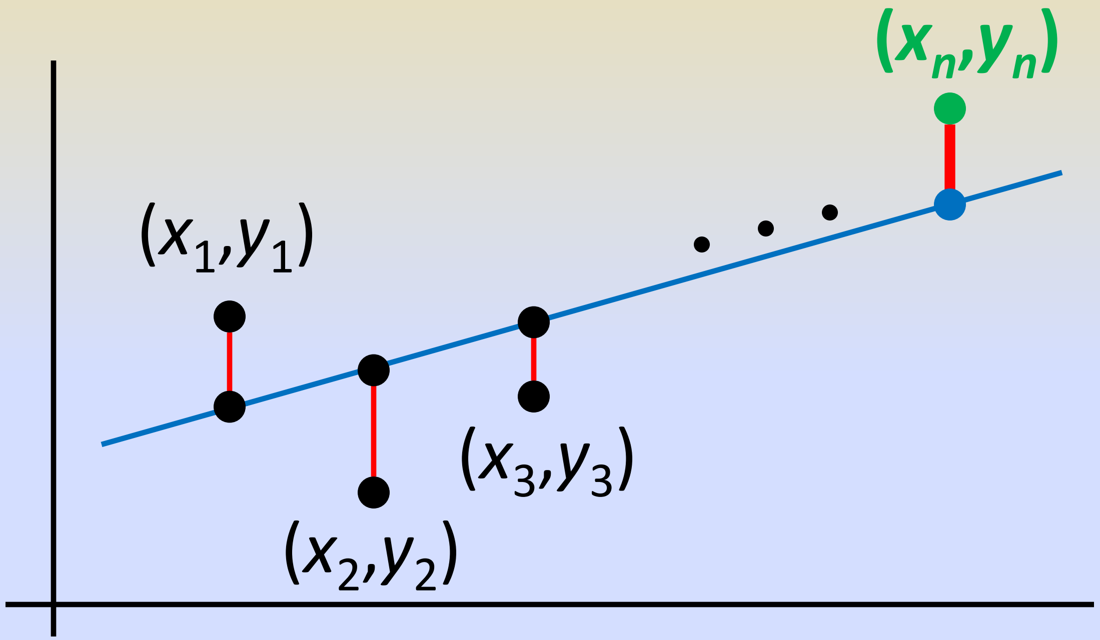
The value of y on the line at x_1 is $y = Ax + B = Ax_1+B$

Best fit line $y = Ax + B$



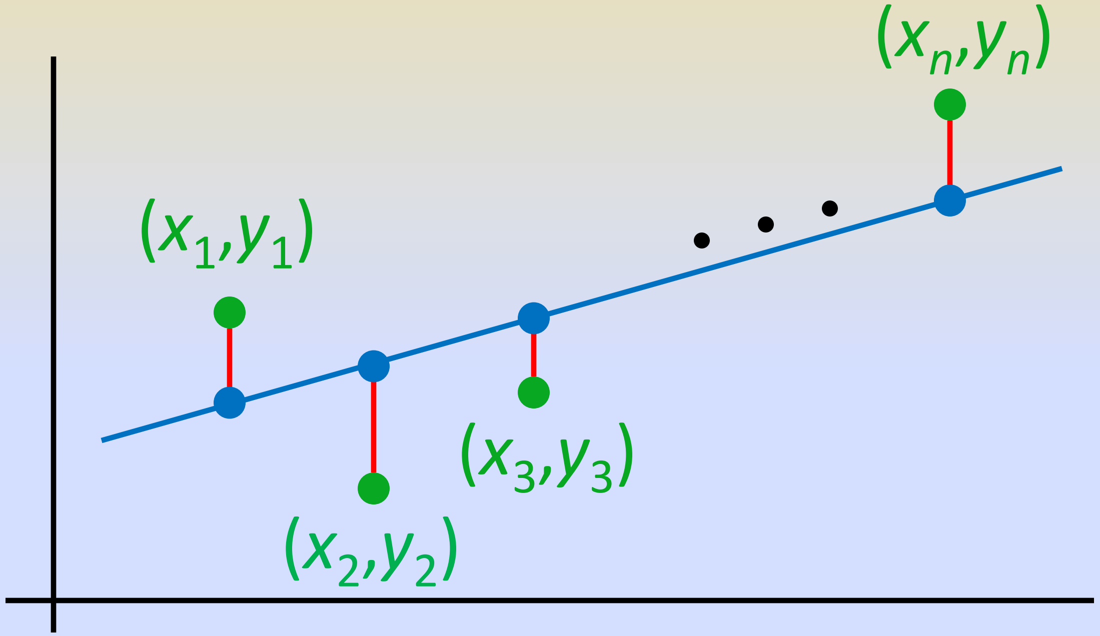
$$E = |Ax_1 + B - y_1| + |Ax_2 + B - y_2| + \ldots + |Ax_n + B - y_n|$$

Best fit line $y = Ax + B$



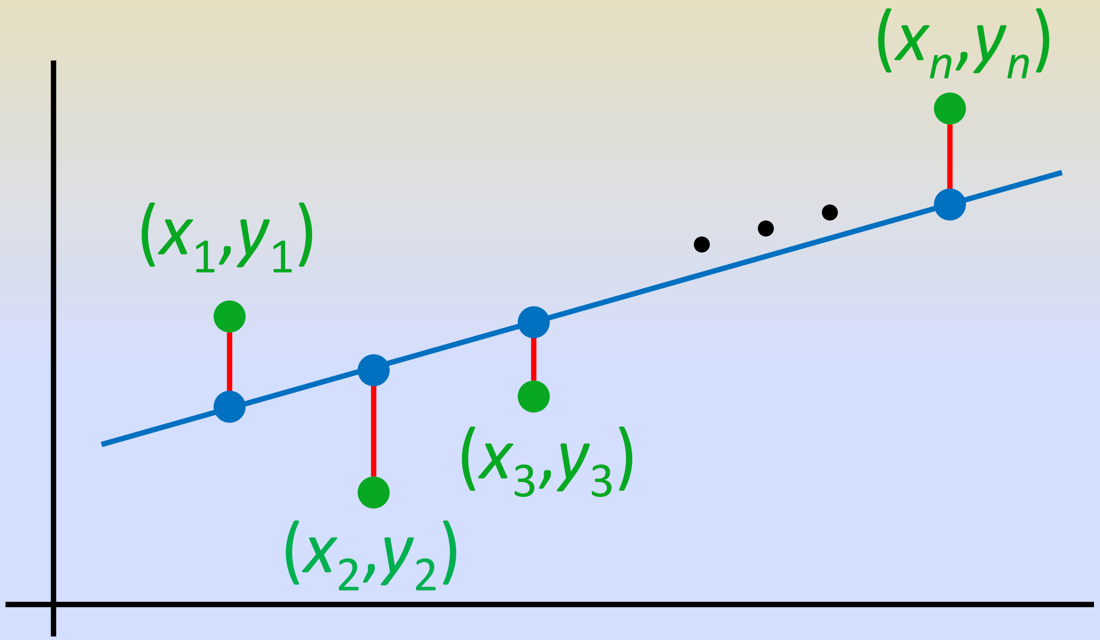
$$E = |Ax_1 + B - y_1| + |Ax_2 + B - y_2| + \ldots + |Ax_n + B - y_n|$$

Best fit line $y = Ax + B$



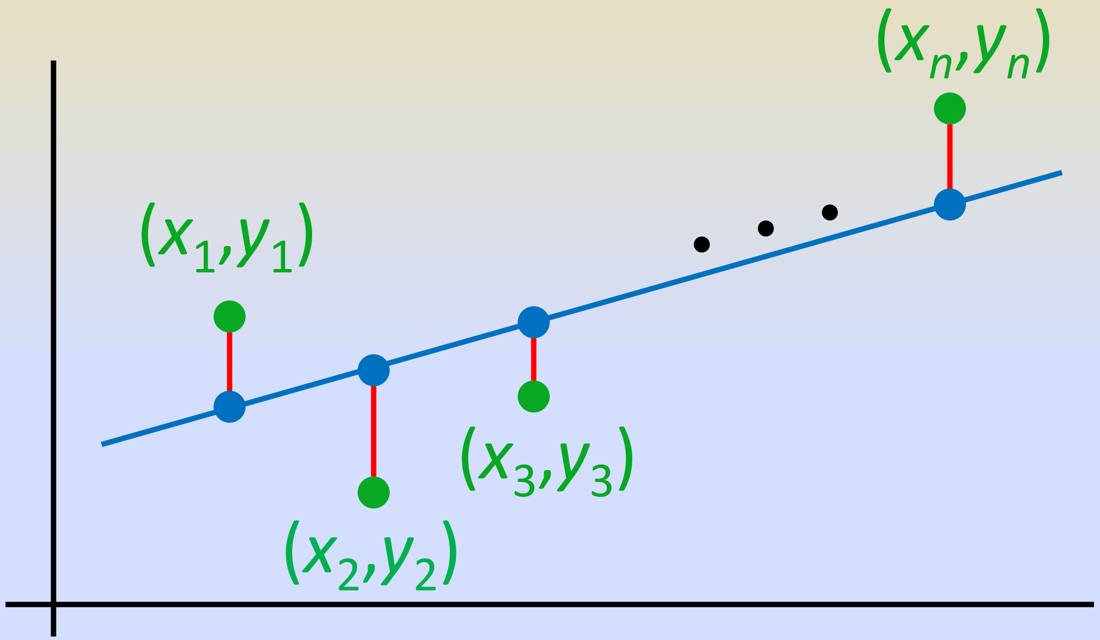
$$E = |Ax_1+B - y_1| + |Ax_2+B - y_2| + \ldots + |Ax_n+B - y_n|$$

Best fit line $y = Ax + B$



$$E = (Ax_1 + B - y_1)^2 + (Ax_2 + B - y_2)^2 + \ldots + (Ax_n + B - y_n)^2$$

Least squares line $y = Ax + B$
(a.k.a. linear regression)



$$E = (Ax_1+B - y_1)^2 + (Ax_2+B - y_2)^2 + \ldots + (Ax_n+B - y_n)^2$$

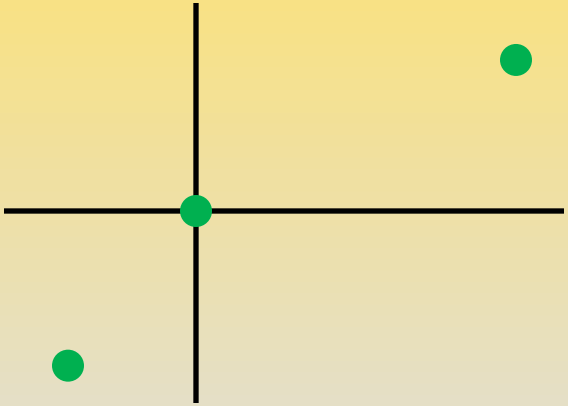
This process is called the method of **least squares**.

Example

x	*y*
-1	-1
0	0
2	1

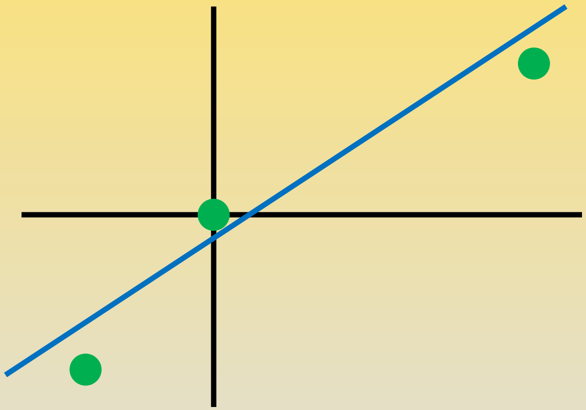
x	*y*
-1	-1
0	0
2	1

Example



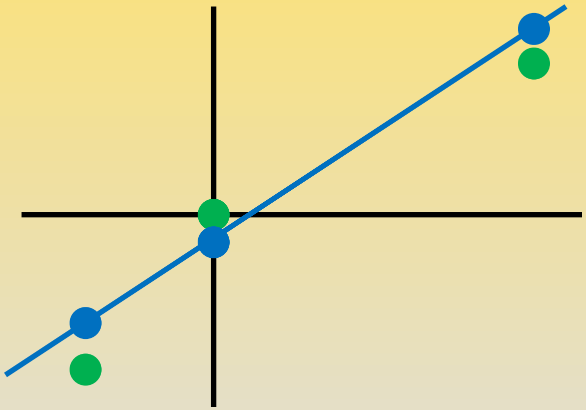
x	*y*
-1	-1
0	0
2	1

Example



x	*y*
-1	-1
0	0
2	1

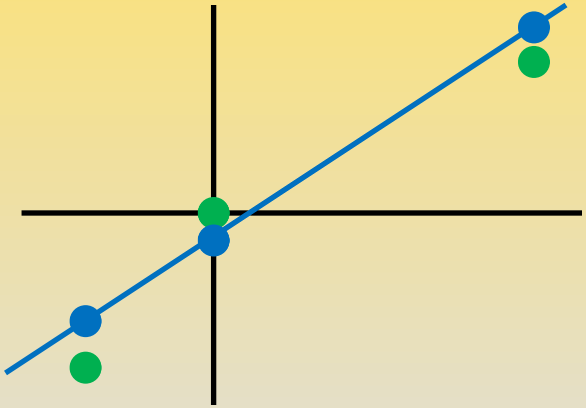
Example



$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

x	y
-1	-1
0	0
2	1

Example

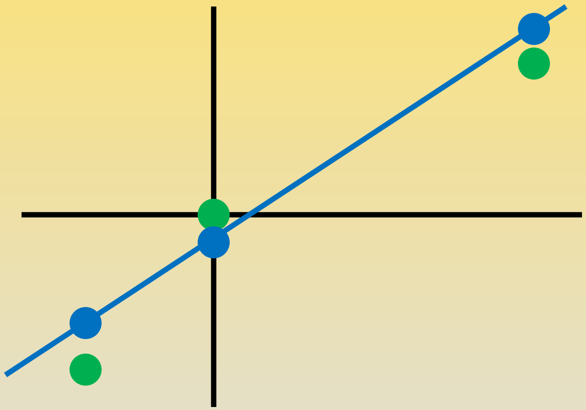


$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

x	y
-1	-1
0	0
2	1

Example



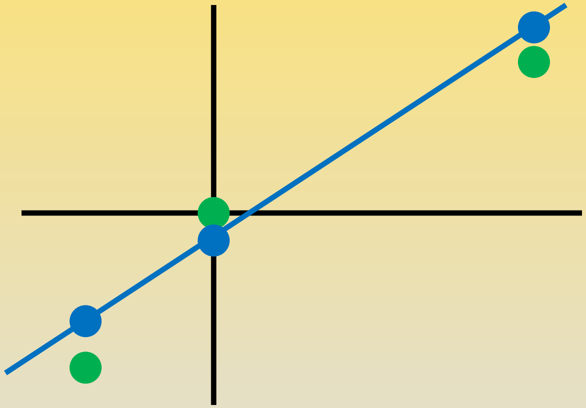
$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

$$\frac{\partial E}{\partial A} = 2(-A + B + 1)^1(-1) + 0 + 2(2A + B - 1)^1(2)$$

x	y
-1	-1
0	0
2	1

Example



$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

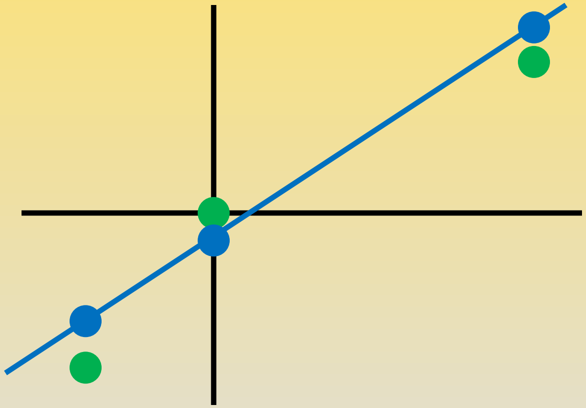
$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

$$\frac{\partial E}{\partial A} = 2(-A+B+1)^1(-1)+0+2(2A+B-1)^1(2)$$

$$= 2A-2B-2+8A+4B-4$$

x	y
-1	-1
0	0
2	1

Example



$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

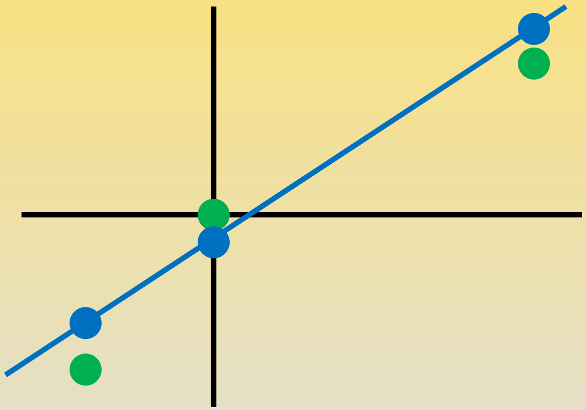
$$\frac{\partial E}{\partial A} = 2(-A+B+1)^1(-1)+0+2(2A+B-1)^1(2)$$

$$= 2A-2B-2+8A+4B-4$$

$$= 10A+2B-6$$

x	y
-1	-1
0	0
2	1

Example



$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

$$\frac{\partial E}{\partial A} = 2(-A+B+1)^1(-1)+0+2(2A+B-1)^1(2)$$

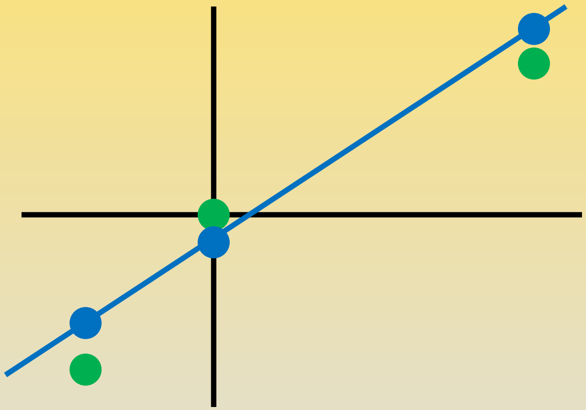
$$= 2A-2B-2+8A+4B-4$$

$$= 10A+2B-6$$

$$\frac{\partial E}{\partial B} = 2(-A+B+1)^1(1)+2B+2(2A+B-1)^1(1)$$

Example

x	y
-1	-1
0	0
2	1



$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

$$\frac{\partial E}{\partial A} = 2(-A+B+1)^1(-1) + 0 + 2(2A+B-1)^1(2)$$

$$= 2A - 2B - 2 + 8A + 4B - 4$$

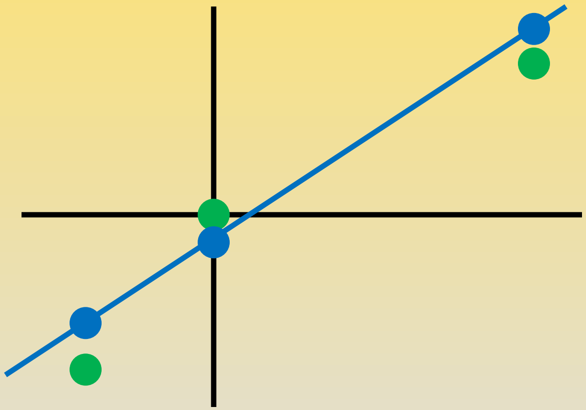
$$= 10A + 2B - 6$$

$$\frac{\partial E}{\partial B} = 2(-A+B+1)^1(1) + 2B + 2(2A+B-1)^1(1)$$

$$= -2A + 2B + 2 + 2B + 4A + 2B - 2$$

Example

x	*y*
-1	-1
0	0
2	1



$$E = (A(-1)+B - (-1))^2 + (A(0)+B - 0)^2 + (A(2)+B - 1)^2$$

$$= (-A + B + 1)^2 + (B)^2 + (2A + B - 1)^2$$

$$\frac{\partial E}{\partial A} = 2(-A+B+1)^1(-1)+0+2(2A+B-1)^1(2)$$

$$= 2A - 2B - 2 + 8A + 4B - 4$$

$$= 10A + 2B - 6$$

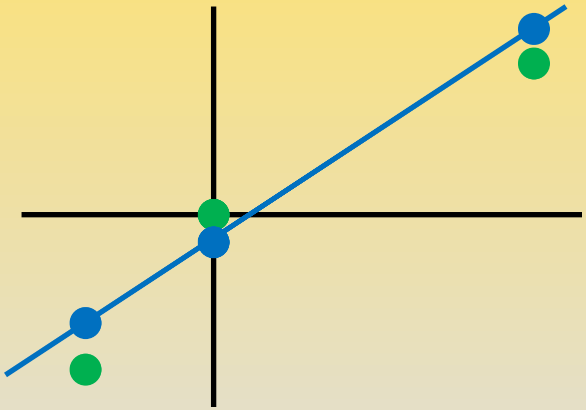
$$\frac{\partial E}{\partial B} = 2(-A+B+1)^1(1)+2B+2(2A+B-1)^1(1)$$

$$= -2A + 2B + 2 + 2B + 4A + 2B - 2$$

$$= 2A + 6B$$

x	*y*
-1	-1
0	0
2	1

Example

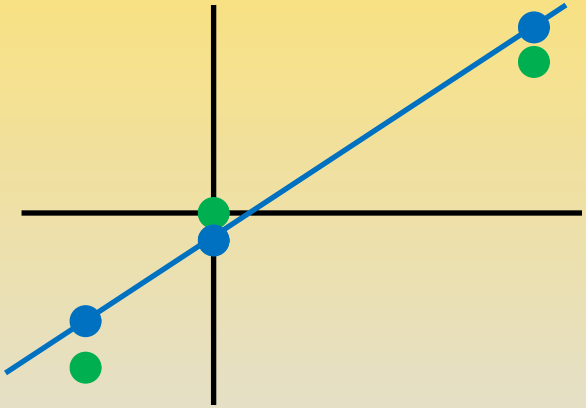


$$\frac{\partial E}{\partial A} = 10A + 2B - 6 = 0$$

$$\frac{\partial E}{\partial B} = 2A + 6B = 0$$

x	*y*
-1	-1
0	0
2	1

Example

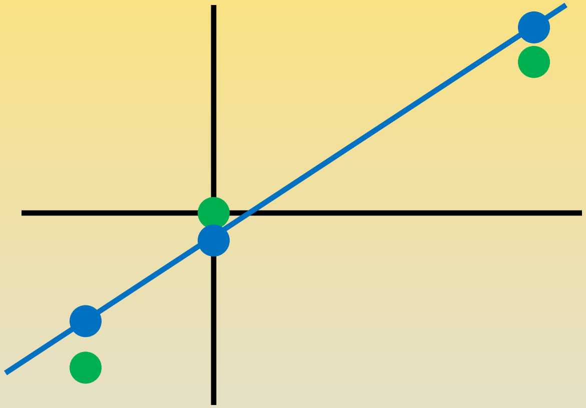


$$\frac{\partial E}{\partial A} = 10A + 2B - 6 = 0 \qquad 10A + 2B = 6$$

$$\frac{\partial E}{\partial B} = 2A + 6B = 0 \qquad 2A + 6B = 0$$

x	y
-1	-1
0	0
2	1

Example



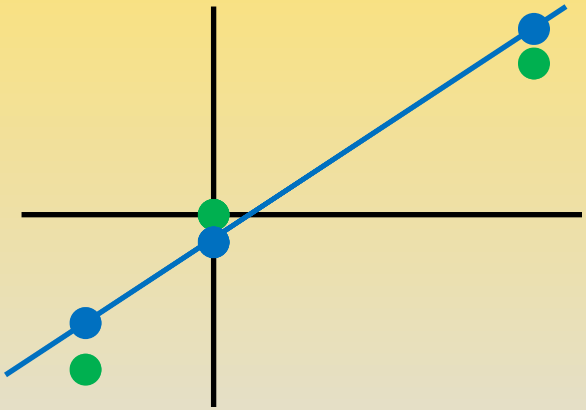
$$\frac{\partial E}{\partial A} = 10A + 2B - 6 = 0 \qquad 10A + 2B = 6$$

$$\frac{\partial E}{\partial B} = 2A + 6B = 0 \qquad 2A + 6B = 0$$

$$A = \frac{9}{14},\ B = -\frac{3}{14} \quad \Rightarrow y = \frac{9}{14}x - \frac{3}{14}$$

x	y
-1	-1
0	0
2	1

Example



$$\frac{\partial E}{\partial A} = 10A + 2B - 6 = 0 \qquad 10A + 2B = 6$$

$$\frac{\partial E}{\partial B} = 2A + 6B = 0 \qquad 2A + 6B = 0$$

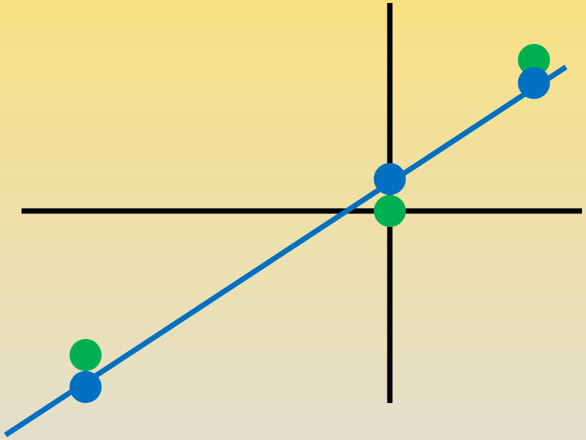
$$A = \frac{9}{14},\ B = -\frac{3}{14} \quad \Rightarrow y = \frac{9}{14}x - \frac{3}{14}$$

Notice:

- The slope is less than 1
- The *y*-intercept is just less than 0.

x	*y*
-2	-1
0	0
1	1

Problem



General formula

- Given the N points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, the best fit line $y = Ax + B$ where

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

and

$$B = \frac{\sum y - A(\sum x)}{N}$$

- Reminder: Σ is “Sigma”, a Greek “S”, for sum.

General formula

- Given the N points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, the best fit line $y = Ax + B$ where

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2} \qquad B = \frac{\sum y - A(\sum x)}{N}$$

$$\sum x = x_1 + x_2 + \cdots + x_N$$

$$\sum y = y_1 + y_2 + \cdots + y_N$$

$$\sum xy = x_1 y_1 + x_2 y_2 + \cdots + x_N y_N$$

$$\sum x^2 = x_1^2 + x_2^2 + \cdots + x_N^2$$

Example

x	*y*		
-1	-1		
0	0		
2	1		

Example

x	y	xy	x^2
-1	-1	1	1
0	0	0	0
2	1	2	4
$\Sigma x = 1$	$\Sigma y = 0$	$\Sigma xy = 3$	$\Sigma x^2 = 5$

Example

x	y	xy	x^2
-1	-1	1	1
0	0	0	0
2	1	2	4
$\Sigma x = 1$	$\Sigma y = 0$	$\Sigma xy = 3$	$\Sigma x^2 = 5$

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2} = \frac{3(3) - (1)(0)}{3(5) - 1^2} = \frac{9}{14}$$

$$B = \frac{\sum y - A(\sum x)}{N} = \frac{0 - \frac{9}{14}(1)}{3} = -\frac{3}{14}$$

Example

x	y	xy	x^2
-1	-1	1	1
0	0	0	0
2	1	2	4
$\Sigma x = 1$	$\Sigma y = 0$	$\Sigma xy = 3$	$\Sigma x^2 = 5$

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2} = \frac{3(3) - (1)(0)}{3(5) - 1^2} = \frac{9}{14}$$

$$B = \frac{\sum y - A(\sum x)}{N} = \frac{0 - \frac{9}{14}(1)}{3} = -\frac{3}{14}$$

$$\text{so } y = \frac{9}{14}x - \frac{3}{14}$$

Problem

x	y	xy	x^2
-2	-1		
0	0		
1	1		
$\Sigma x =$	$\Sigma y =$	$\Sigma xy =$	$\Sigma x^2 =$

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2} =$$

$$B = \frac{\sum y - A(\sum x)}{N} =$$

Reminder

- Given the N points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, the (least squares) line that bests fit the data is $y = Ax + B$ where

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

and

$$B = \frac{\sum y - A(\sum x)}{N} \quad .$$

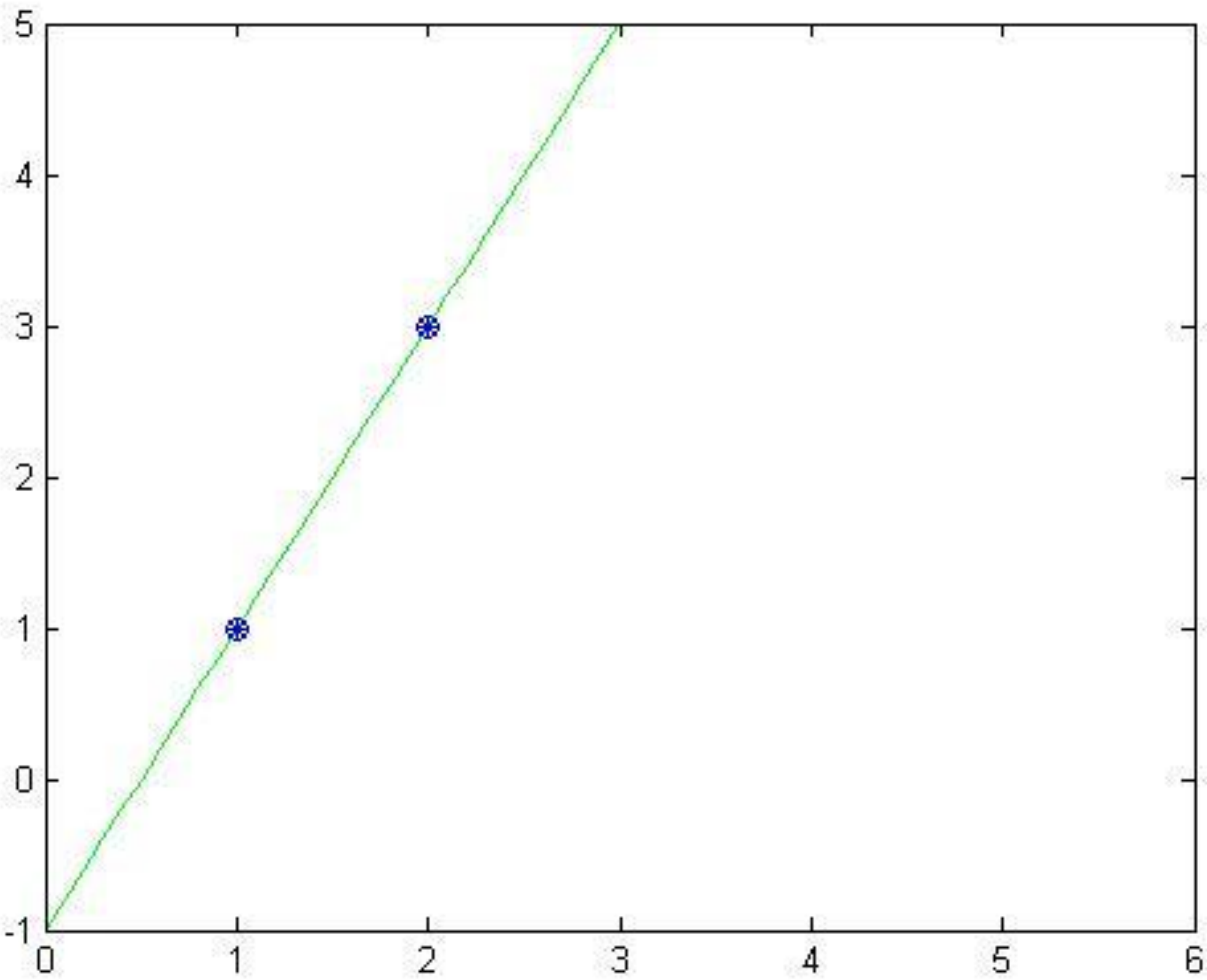
Example

- Find best fit lines using either the first 2, 3, 4 or all 5 points of

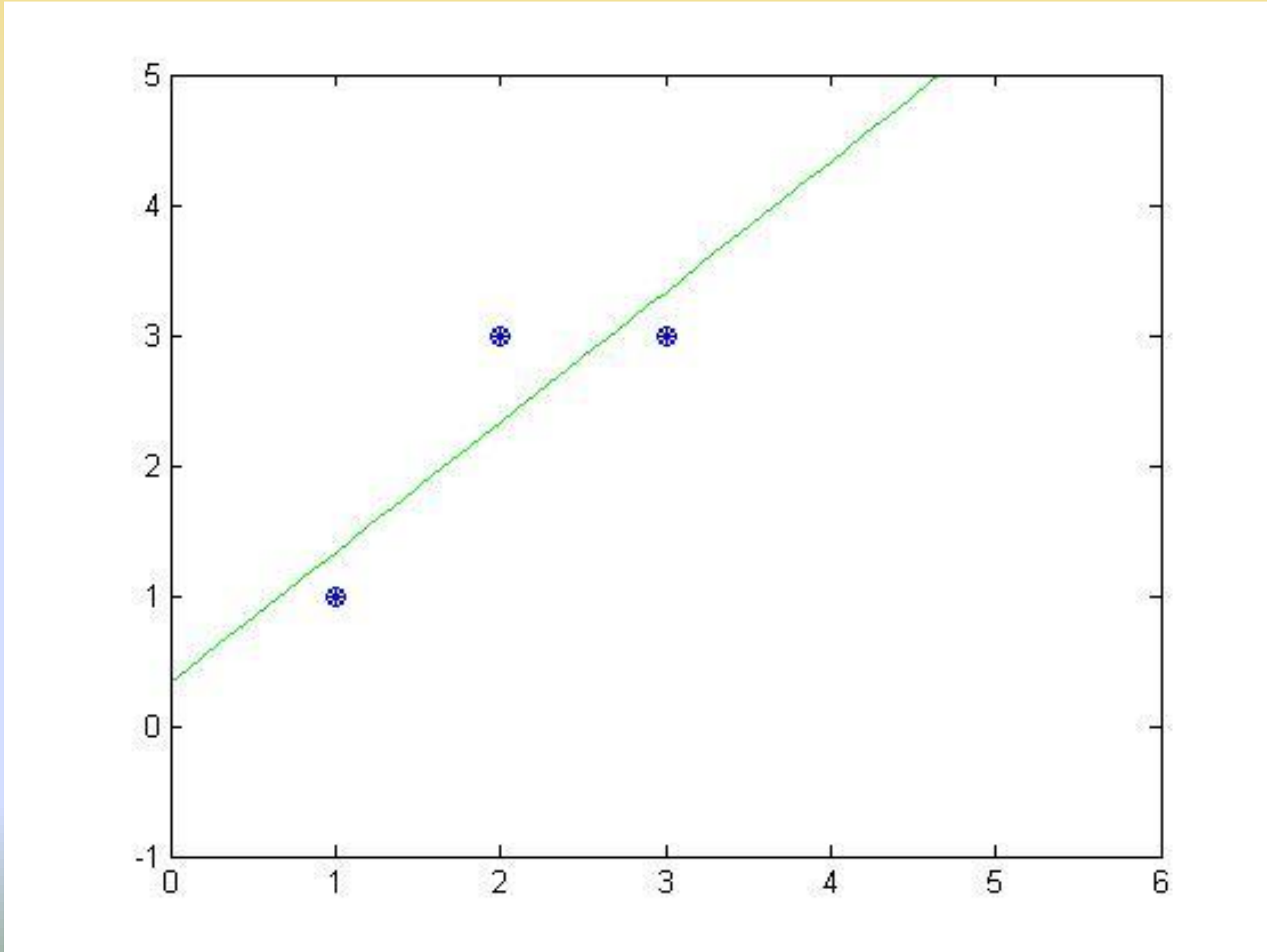
x	*y*	*xy*	x^2
1	1		
2	3		
3	3		
4	4		
5	0		

Points	Line
2	
3	
4	
5	

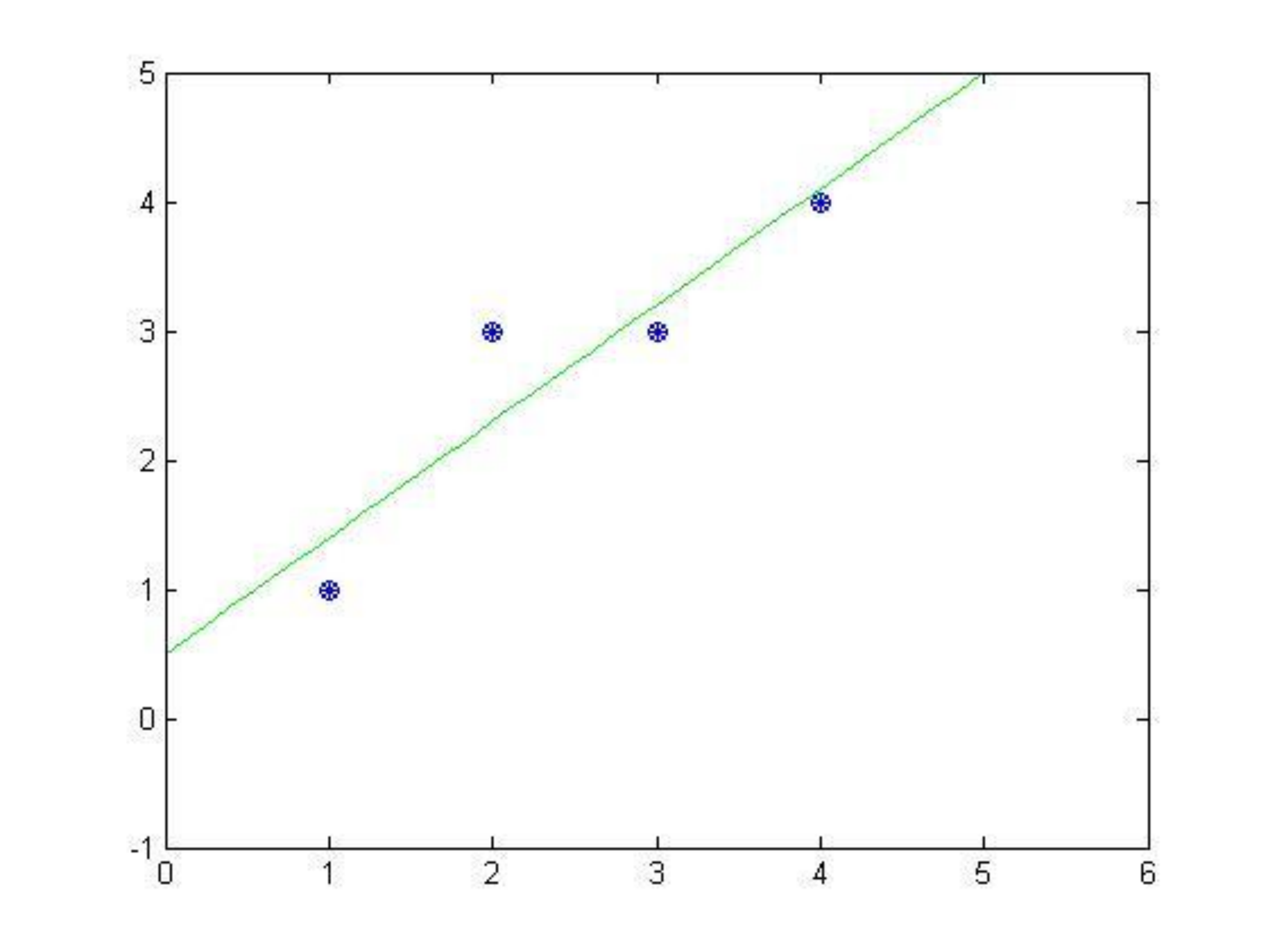
$$y = 2x - 1$$



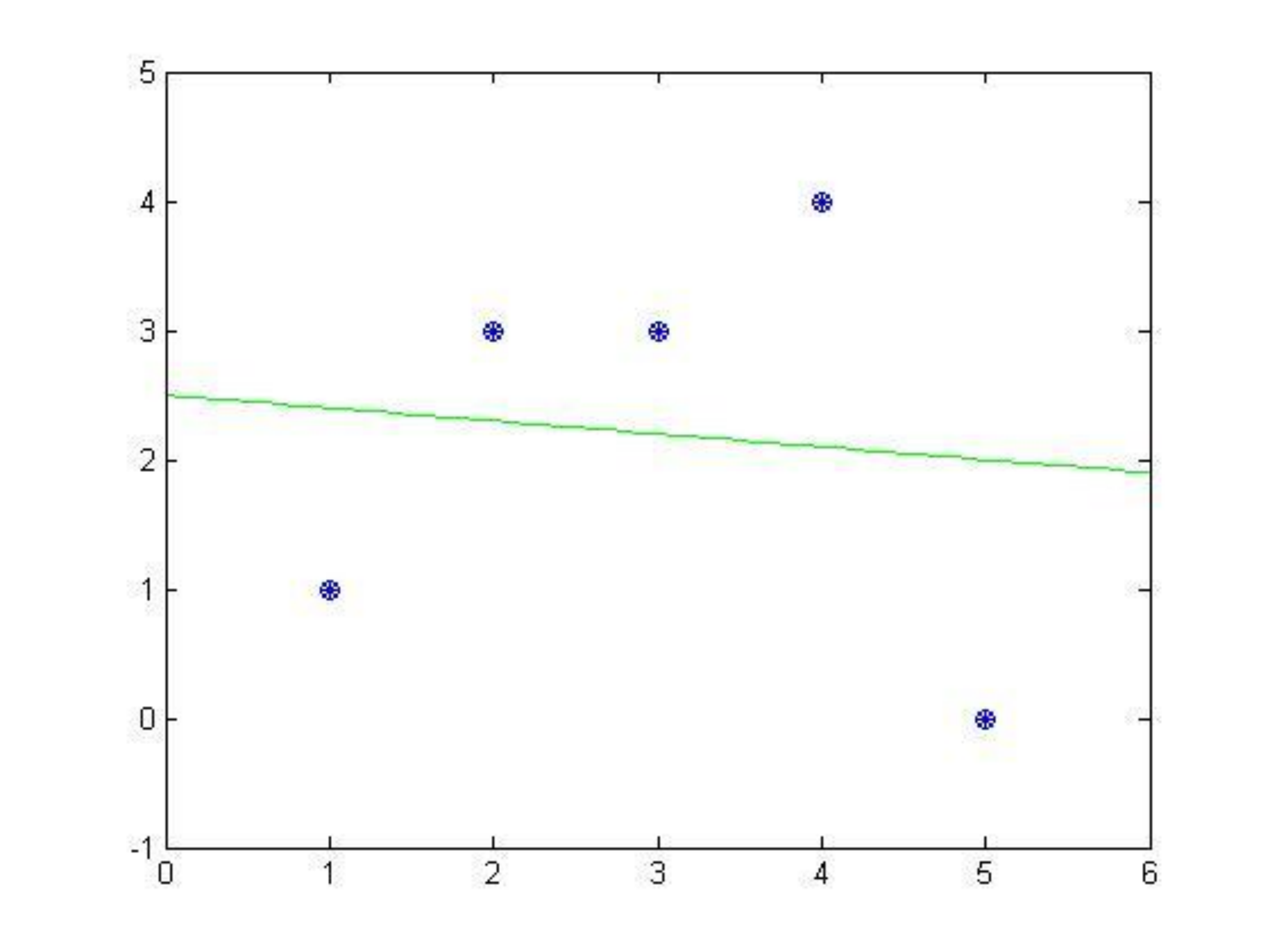
$$y = \qquad x + \frac{1}{3}$$



$$y = \frac{9}{10}x + \frac{1}{2}$$



$$y = -\frac{1}{10}x + \frac{5}{2}$$



Reminder

- Given the N points $(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$, the (least squares) line that bests fit the data is $y = Ax + B$ where

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

and

$$B = \frac{\sum y - A(\sum x)}{N} \quad .$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1) \cdot x_1 + \cdots + 2(Ax_N + B - y_N) \cdot x_N$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1)\cdot x_1 + \cdots + 2(Ax_N + B - y_N)\cdot x_N$$

$$= 2A(x_1^2 + \cdots + x_N^2) + 2B(x_1 + \cdots + x_N) - 2(x_1 y_1 + \cdots + x_N y_N)$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1)\cdot x_1 + \cdots + 2(Ax_N + B - y_N)\cdot x_N$$

$$= 2A(x_1^2 + \cdots + x_N^2) + 2B(x_1 + \cdots + x_N) - 2(x_1 y_1 + \cdots + x_N y_N)$$

$$= 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy)$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1)\cdot x_1 + \cdots + 2(Ax_N + B - y_N)\cdot x_N$$

$$= 2A(x_1^2 + \cdots + x_N^2) + 2B(x_1 + \cdots + x_N) - 2(x_1 y_1 + \cdots + x_N y_N)$$

$$= 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy)$$

$$\frac{\partial E}{\partial B} = 2(Ax_1 + B - y_1)\cdot 1 + \cdots + 2(Ax_N + B - y_N)\cdot 1$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1)\cdot x_1 + \cdots + 2(Ax_N + B - y_N)\cdot x_N$$

$$= 2A(x_1^2 + \cdots + x_N^2) + 2B(x_1 + \cdots + x_N) - 2(x_1 y_1 + \cdots + x_N y_N)$$

$$= 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy)$$

$$\frac{\partial E}{\partial B} = 2(Ax_1 + B - y_1)\cdot 1 + \cdots + 2(Ax_N + B - y_N)\cdot 1$$

$$= 2A(x_1 + \cdots + x_N) + 2B(1 + \cdots + 1) - 2(y_1 + \cdots + y_N)$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1)\cdot x_1 + \cdots + 2(Ax_N + B - y_N)\cdot x_N$$

$$= 2A(x_1^2 + \cdots + x_N^2) + 2B(x_1 + \cdots + x_N) - 2(x_1 y_1 + \cdots + x_N y_N)$$

$$= 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy)$$

$$\frac{\partial E}{\partial B} = 2(Ax_1 + B - y_1)\cdot 1 + \cdots + 2(Ax_N + B - y_N)\cdot 1$$

$$= 2A(x_1 + \cdots + x_N) + 2B(1 + \cdots + 1) - 2(y_1 + \cdots + y_N)$$

$$= 2A(\sum x) + 2B(\sum 1) - 2(\sum y)$$

Derivation of the formulas

$$E = (Ax_1 + B - y_1)^2 + \cdots + (Ax_N + B - y_N)^2$$

$$\frac{\partial E}{\partial A} = 2(Ax_1 + B - y_1)\cdot x_1 + \cdots + 2(Ax_N + B - y_N)\cdot x_N$$

$$= 2A(x_1^2 + \cdots + x_N^2) + 2B(x_1 + \cdots + x_N) - 2(x_1 y_1 + \cdots + x_N y_N)$$

$$= 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy)$$

$$\frac{\partial E}{\partial B} = 2(Ax_1 + B - y_1)\cdot 1 + \cdots + 2(Ax_N + B - y_N)\cdot 1$$

$$= 2A(x_1 + \cdots + x_N) + 2B(1 + \cdots + 1) - 2(y_1 + \cdots + y_N)$$

$$= 2A(\sum x) + 2B(\ N\ ) - 2(\sum y)$$

Derivation of the formulas

$$\frac{\partial E}{\partial A} = 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy)$$

$$\frac{\partial E}{\partial B} = 2A(\sum x) + 2B(\ N\ ) - 2(\sum y)$$

Derivation of the formulas

$$\frac{\partial E}{\partial A} = 2A(\sum x^2) + 2B(\sum x) - 2(\sum xy) \quad = 0$$

$$\frac{\partial E}{\partial B} = 2A(\sum x) + 2B(\ N\ ) - 2(\sum y) \quad = 0$$

Derivation of the formulas

$$2A\left(\sum x^2\right) + 2B\left(\sum x\right) - 2\left(\sum xy\right) = 0$$

$$2A\left(\sum x\right) + 2B(\ N\ ) - 2\left(\sum y\right) = 0$$

Derivation of the formulas

$$2A(\sum x^2) + 2B(\sum x) = 2(\sum xy)$$

$$2A(\sum x) + 2B(\ N\ ) = 2(\sum y)$$

Derivation of the formulas

$$A\left(\sum x^2\right) + B\left(\sum x\right) = \sum xy$$

$$A\left(\sum x\right) + B\left(N\right) = \sum y$$

Solve for *A* and *B*

$$A(\sum x^2) + B(\sum x) = \sum xy$$

$$A(\sum x \ \ ) + B(\ N\ ) = \sum y$$

Solve for *A* and *B*

$$(\sum x^2)A + (\sum x)B = \sum xy$$

$$(\sum x\ )A + (\ N\ )B = \sum y$$

Solve for *A* and *B*

$$(\sum x^2)A + (\sum x)B = \sum xy$$

$$(\sum x)A + (N)B = \sum y$$

$$\begin{bmatrix} \sum x^2 & \sum x \\ \sum x & N \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

Solve for *A* and *B*

$$(\sum x^2)A + (\sum x)B = \sum xy$$

$$(\sum x \ \ )A + (\ N\ )B = \sum y$$

$$\begin{bmatrix} \sum x^2 & \sum x \\ \sum x & N \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

$$\begin{bmatrix} \sum x^2 & \sum x \\ \sum x & N \end{bmatrix}^{-1} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix}$$

Solve for *A* and *B*

$$(\sum x^2)A + (\sum x)B = \sum xy$$

$$(\sum x\ \ )A + (\ N\ )B = \sum y$$

$$\begin{bmatrix} \sum x^2 & \sum x \\ \sum x & N \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

$$\begin{bmatrix} \sum x^2 & \sum x \\ \sum x & N \end{bmatrix}^{-1} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix}$$

so $$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

Solve for *A* and *B*

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

Solve for *A* and *B*

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

Solve for *A* and *B*

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

$$= \frac{1}{N \cdot \sum x^2 - (\sum x)^2} \begin{bmatrix} N(\sum xy) + (-\sum x)(\sum y) \\ (-\sum x)(\sum xy) + (\sum x^2)(\sum y) \end{bmatrix}$$

Solve for *A* and *B*

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

$$= \frac{1}{N \cdot \sum x^2 - (\sum x)^2} \begin{bmatrix} N(\sum xy) + (-\sum x)(\sum y) \\ (-\sum x)(\sum xy) + (\sum x^2)(\sum y) \end{bmatrix}$$

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

Solve for *A* and *B*

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

$$= \frac{1}{N \cdot \sum x^2 - (\sum x)^2} \begin{bmatrix} N(\sum xy) + (-\sum x)(\sum y) \\ (-\sum x)(\sum xy) + (\sum x^2)(\sum y) \end{bmatrix}$$

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

$$B = \frac{(-\sum x)(\sum xy) - (\sum x^2)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

Solve for *A* and *B*

$$\begin{bmatrix} A \\ B \end{bmatrix} = \frac{1}{N(\sum x^2) - (\sum x)^2} \begin{bmatrix} N & -\sum x \\ -\sum x & \sum x^2 \end{bmatrix} \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

$$= \frac{1}{N \cdot \sum x^2 - (\sum x)^2} \begin{bmatrix} N(\sum xy) + (-\sum x)(\sum y) \\ (-\sum x)(\sum xy) + (\sum x^2)(\sum y) \end{bmatrix}$$

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

$$B = \frac{(-\sum x)(\sum xy) - (\sum x^2)(\sum y)}{N(\sum x^2) - (\sum x)^2} = \frac{\sum y - A(\sum x)}{N}$$

Example 2 in Section 7.5
Lung cancer deaths as a result of smoking

Country	Cigarettes	Lung Cancer Deaths
Norway	250	95
Sweden	300	120
Denmark	350	165
Australia	470	170

Example 2 in Section 7.5
Lung cancer deaths as a result of smoking

Country	Cigarettes	Lung Cancer Deaths
Norway	250	95
Sweden	300	120
Denmark	350	165
Australia	470	170

Best fit line: $y = .338x + 21.621$

y ↑ Deaths x ↑ Cigs

Example 2 in Section 7.5
Lung cancer deaths as a result of smoking

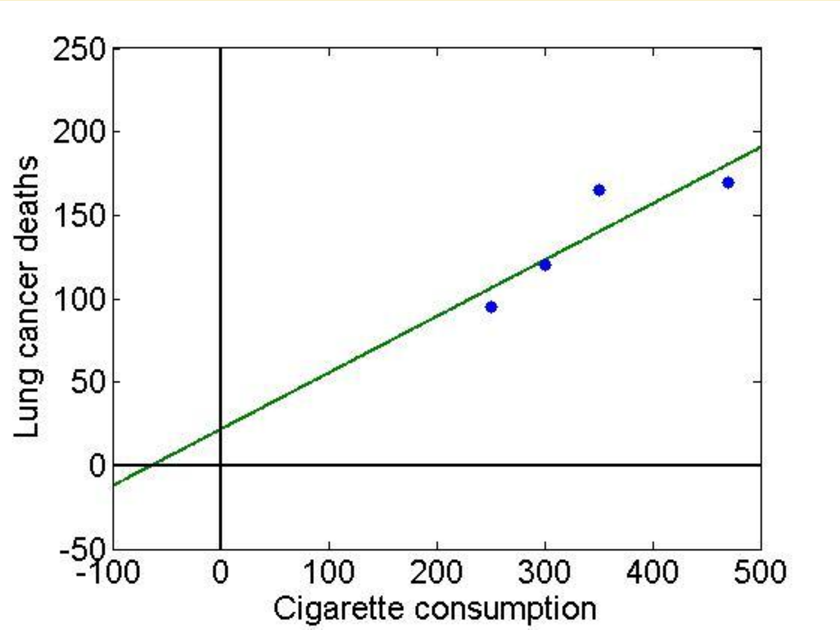
Country	Cigarettes	Lung Cancer Deaths
Norway	250	95
Sweden	300	120
Denmark	350	165
Australia	470	170

Best fit line: $y = .338x + 21.621$

y ↑ Deaths; x ↑ Cigs

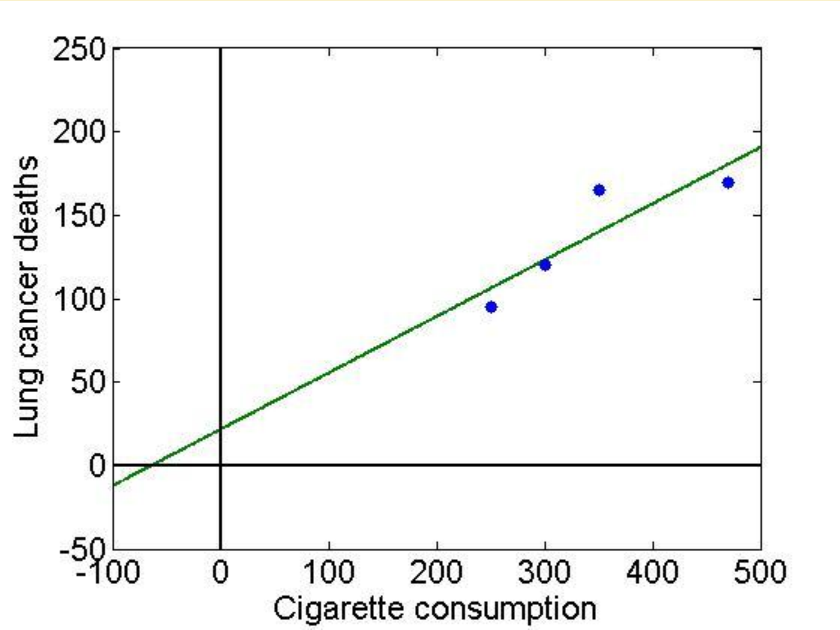
What do .338 and 21.621 represent?

$y = .338x + 21.621$



Suppose a country estimates that its cigarette consumption is 400 per person per year.
What lung cancer deaths can it expect?

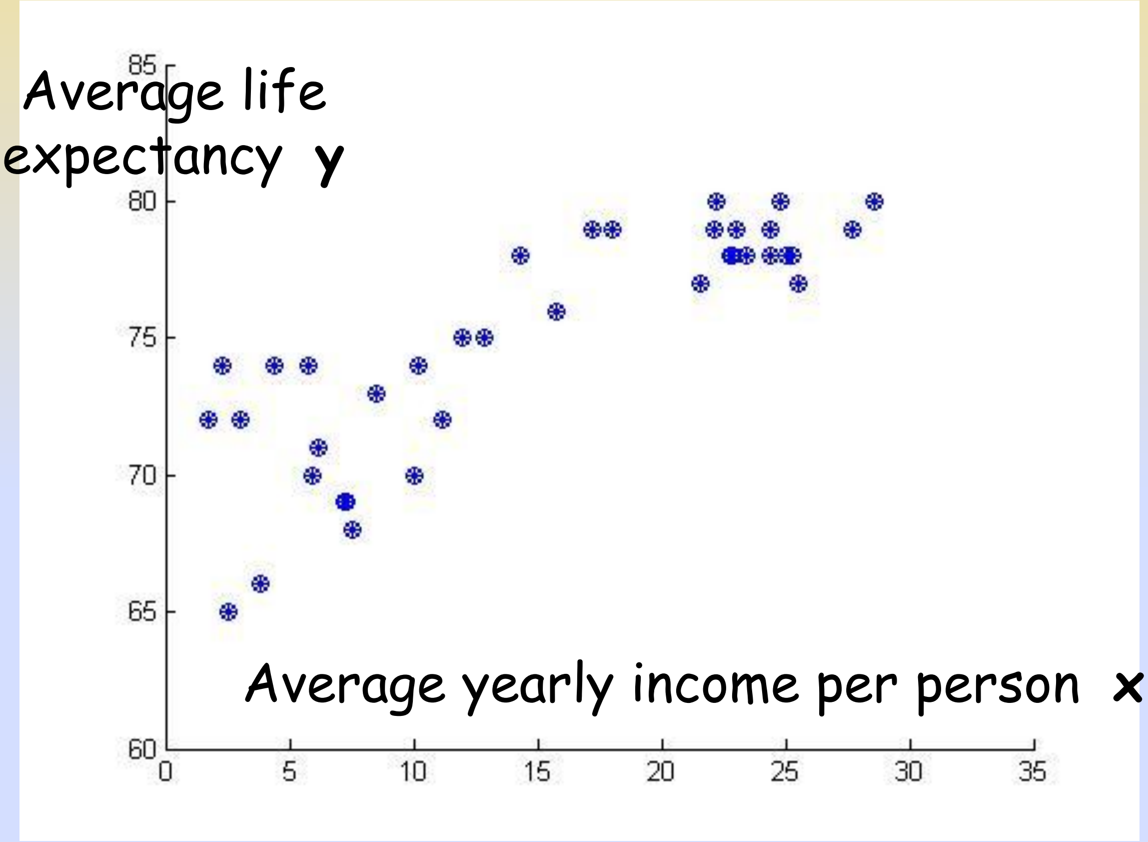
$$y = .338x + 21.621$$



Suppose a country estimates that its cigarette consumption is 400 per person per year.
What lung cancer deaths can it expect?

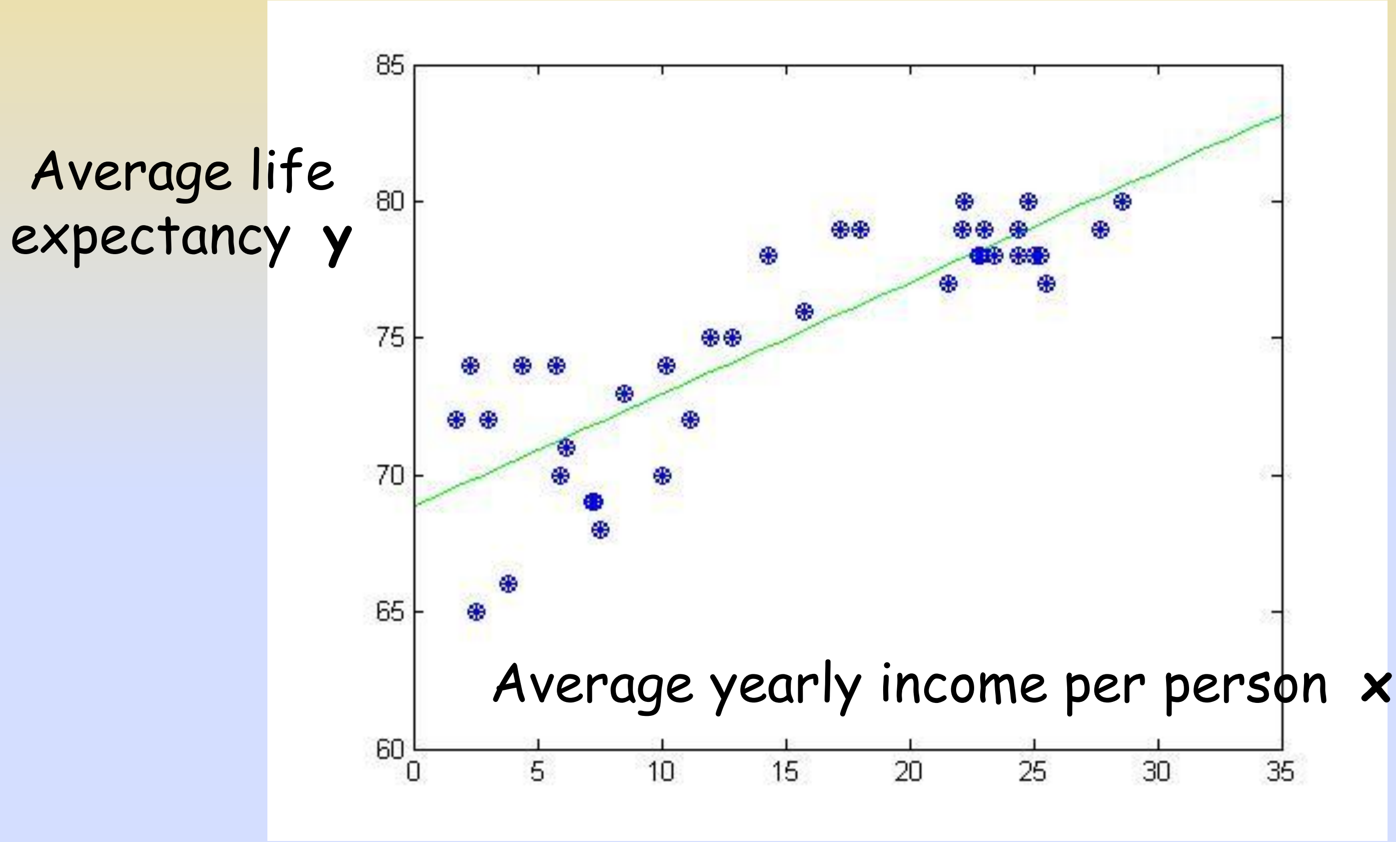
Suppose a country wants to reduce its lung cancer death rate to 50 deaths (per million males per year).
What do they need to reduce the smoking rate to?

Average life expectancy as a function of income.



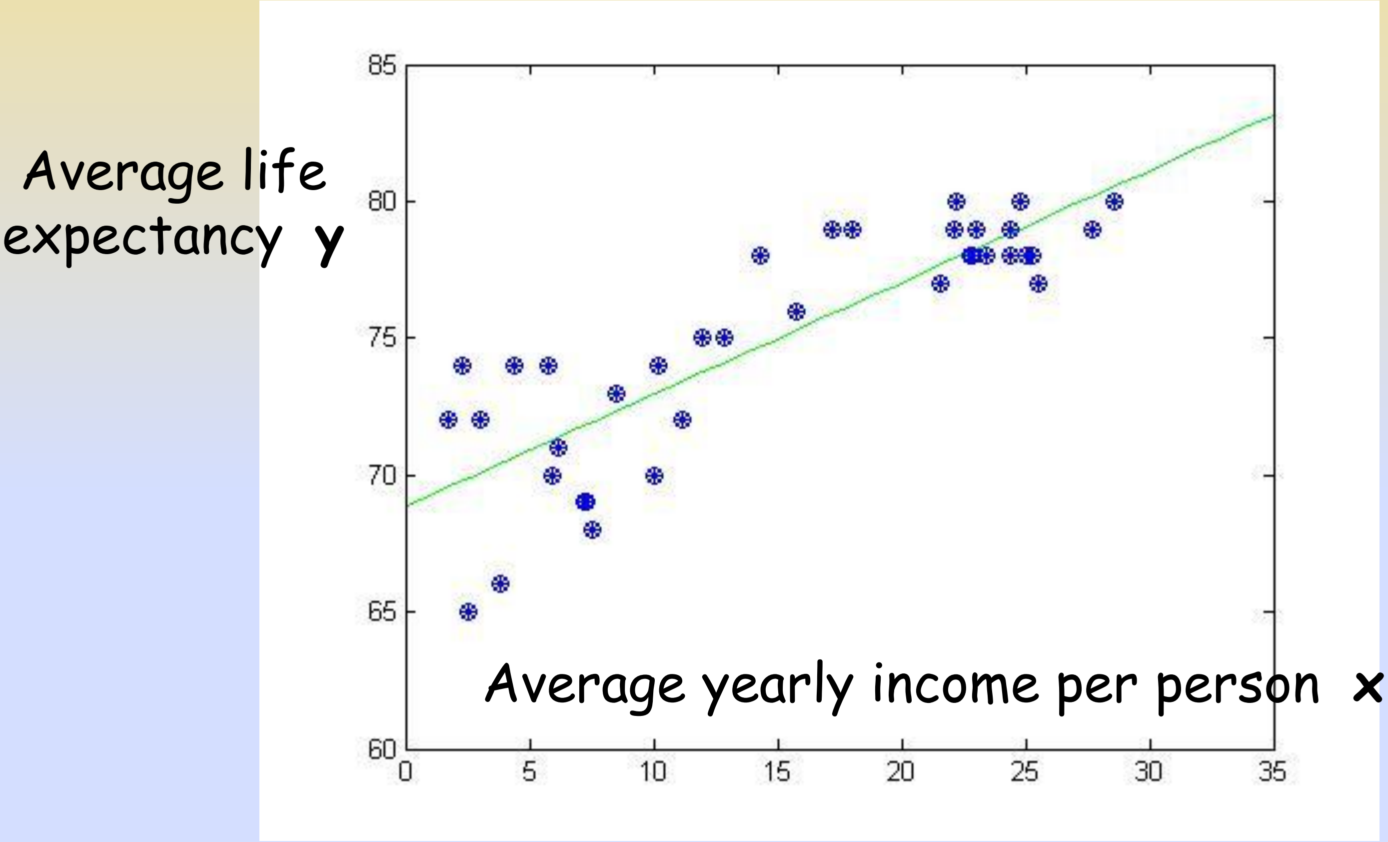
$$y = .4085x + 68.85$$

Age (y) ← → Income (x)



$$y = .4085x + 68.85$$

Age ↑ (y) Income ↑ (x)



What do .4085 and 68.85 represent?

Main points

- We can find the line, $y = Ax + B$, that best fits any given data:

$$A = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2} \qquad B = \frac{y - A(\sum x)}{N}$$

- Meaning of slope *A* and *y*-intercept *B.*
- Usefulness of line:
  - Given a particular value of *x*, find *y*.
  - Given a particular value of *y*, find *x*.