Regression Line

Gaston Sanchez

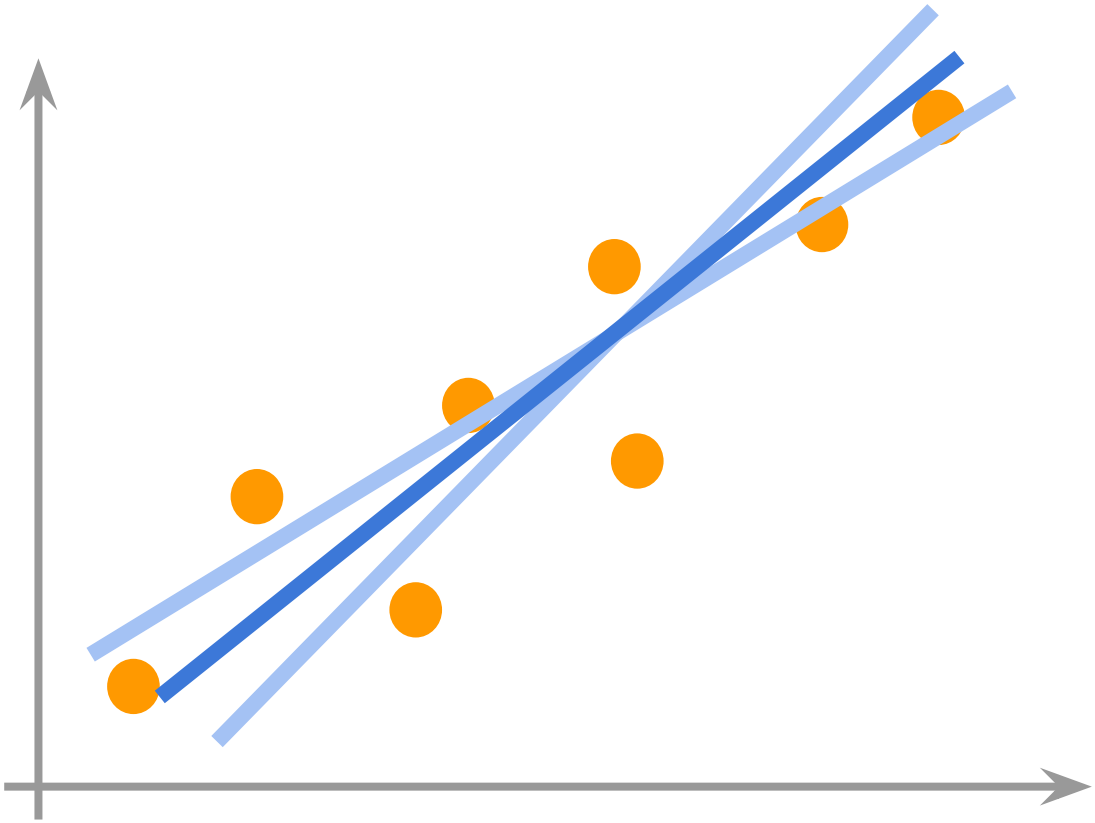
Creative Commons Attribution Share-Alike 4.0 International CC BY-SA

Regression Line

Fitting a line between X and Y

Fitting a line

which is the best line?



How to find the “best” fitting line?

Line formula

$$y = mx + b$$

Line formula

how much the y-value changes when
the x-value increases by 1 unit

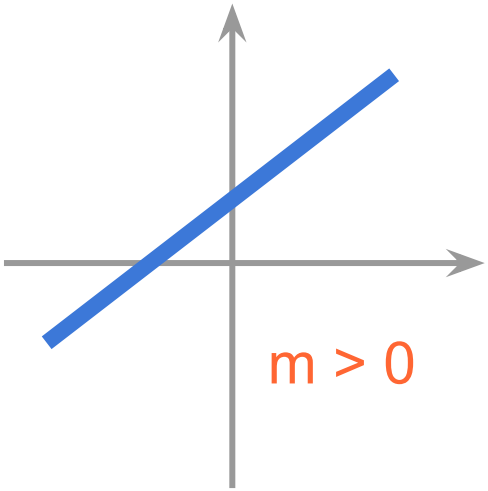
slope

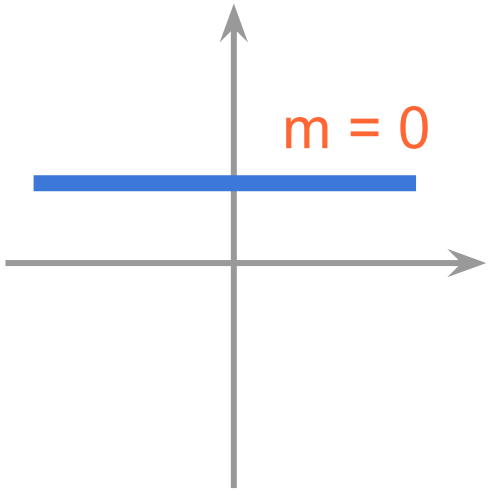
$$y = mx + b$$

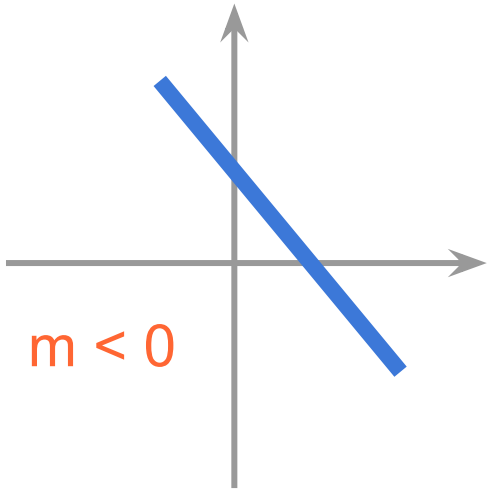
intercept

y-value of the point of interception
of the line and the y-axis

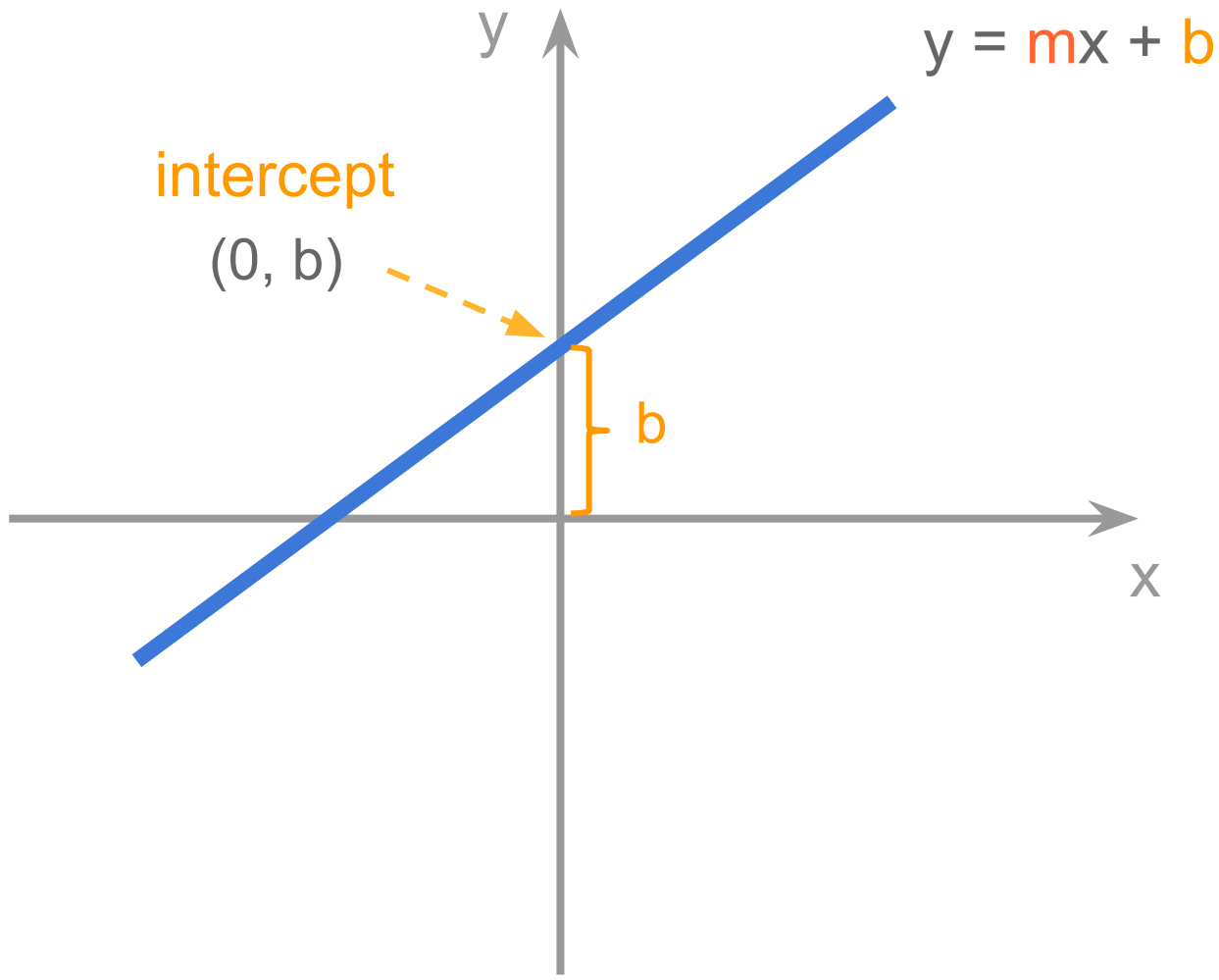
Types of slopes



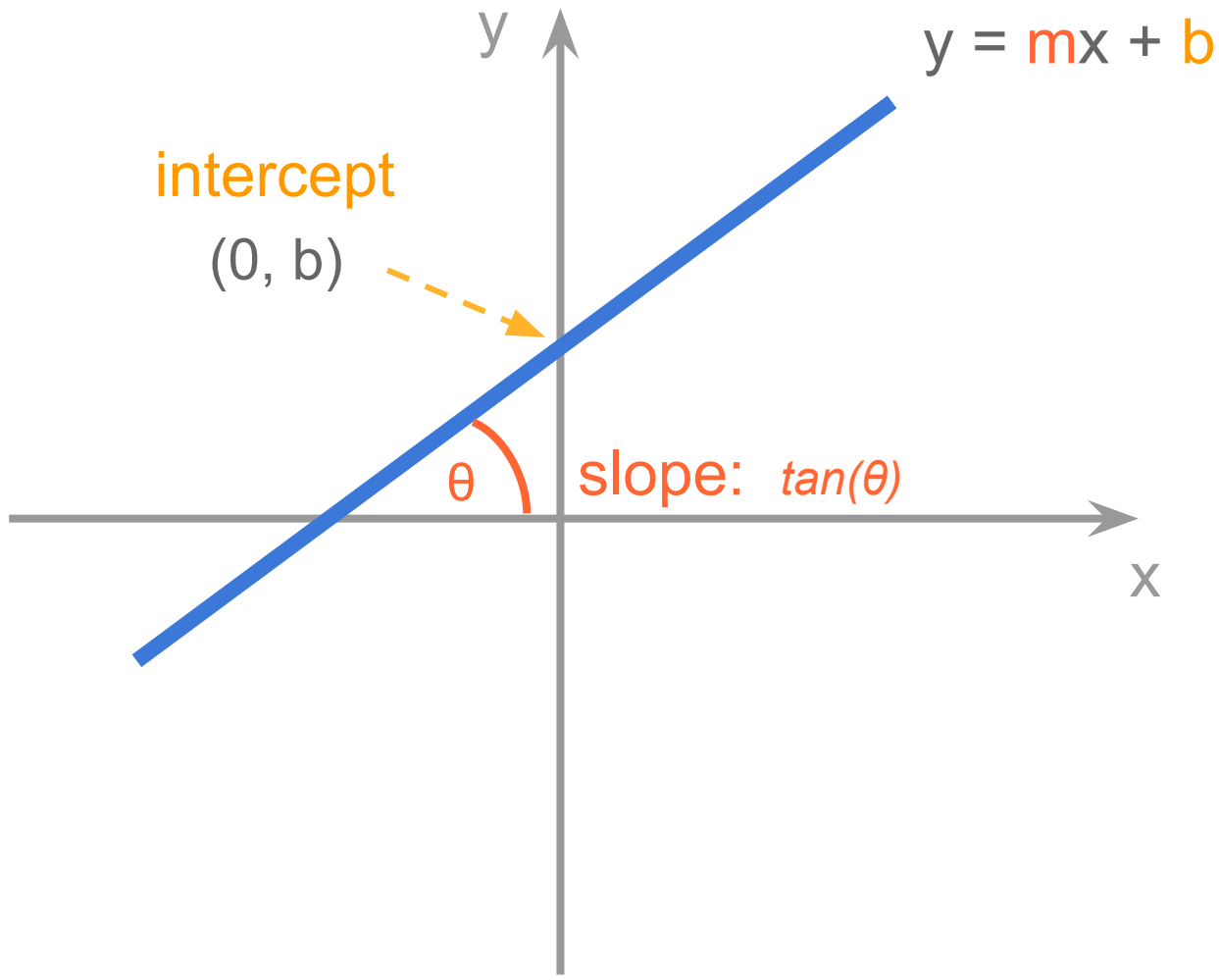




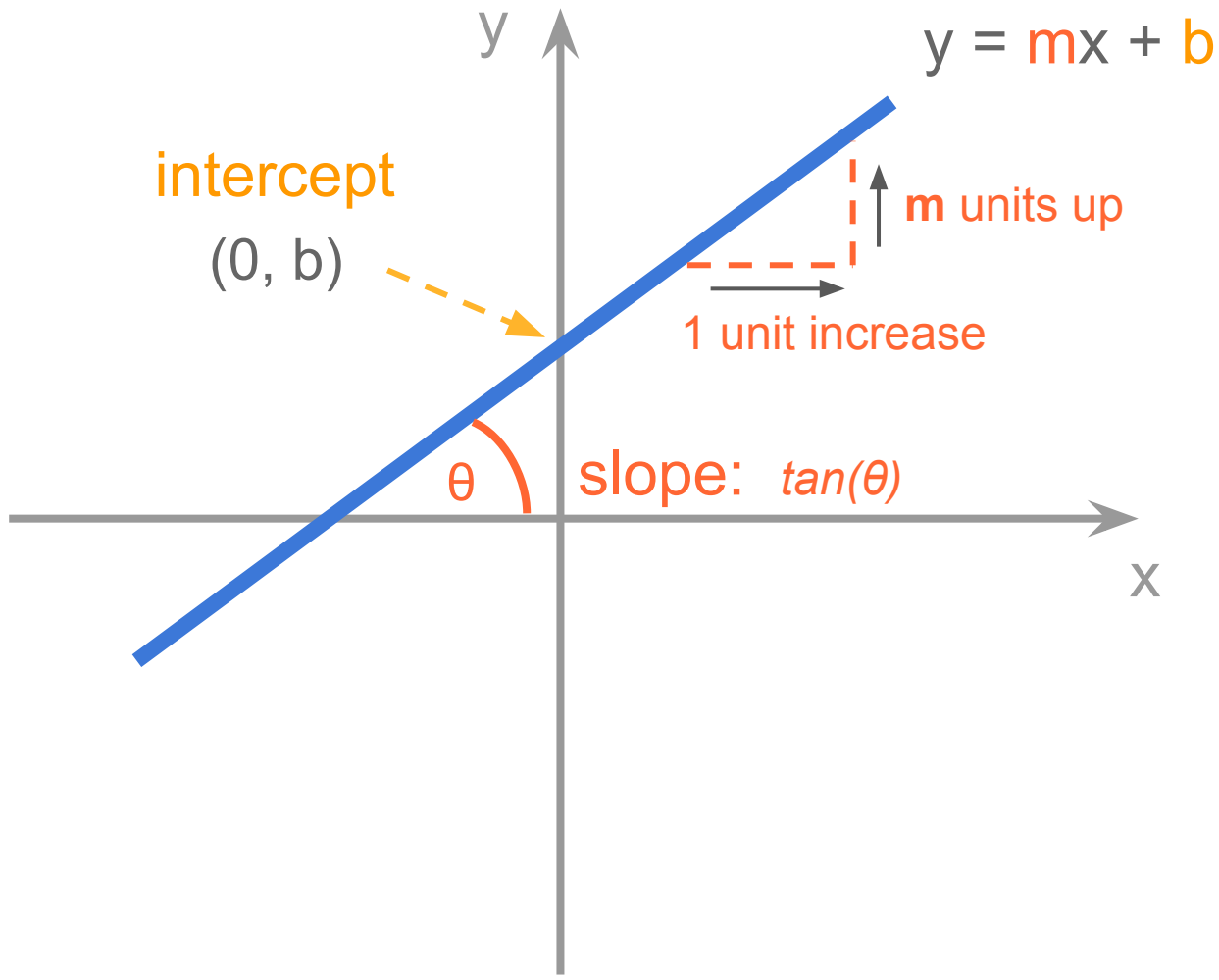
Line formula



Line formula



Line formula

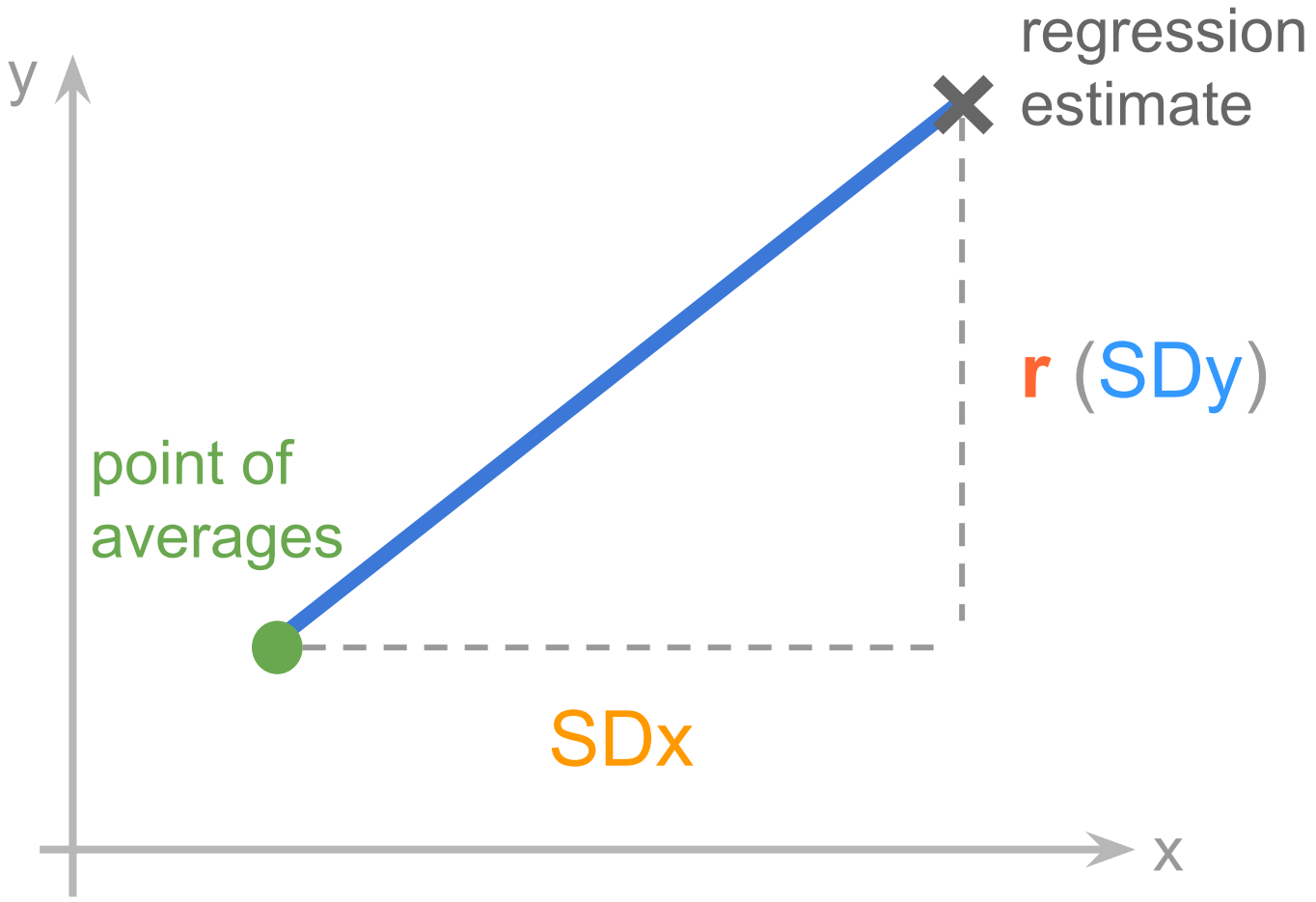


Regression Line formula

$$\textit{predicted}\ y = mx + b$$

Reminder: Associated with each increase of one SD in **x** there is an increase of only **r** SDs in **y**, on the average

Regression Method



Regression Line formula

$$\mathbf{m} = r \left(\frac{SDy}{SDx} \right)$$

$$\mathbf{b} = mean(y) - r \left(\frac{SDy}{SDx} \right) mean(x)$$

Finding the regression line

Group of men with:

X: avg height = **70** inches, SD = **3** inches

Y: avg weight = **180** pounds, SD = **45** pounds

- r = 0.40

$$\mathbf{m} = 0.40 \left(\frac{45}{3} \right) = 6$$

Finding the regression line

Group of men with:

X: avg height = **70** inches, SD = **3** inches

Y: avg weight = **180** pounds, SD = **45** pounds

- r = 0.40

$$\mathbf{b} = 180 - 0.4 \left(\frac{45}{3} \right) 70 = -240$$

Finding the regression line

Group of men with:

X: avg height = **70** inches, SD = **3** inches

Y: avg weight = **180** pounds, SD = **45** pounds

- r = 0.40

$$\textit{predicted}\ y = 6x - 240$$

Finding the regression line

Predict weight for a man with height 60 inches

$$\textit{predicted}\ y = 6x - 240$$

$$\textit{pred}\ y = 6\ (60) - 240 = 120 \text{ pounds}$$

Finding the regression line

Predict weight for a man with height 78 inches

$$\textit{predicted}\ y = 6x - 240$$

$$\textit{pred}\ y = 6\ (78) - 240 = 228 \text{ pounds}$$

Finding the regression line

Predict weight for a man with height 0 inches

$$\textit{predicted}\ y = 6x - 240$$

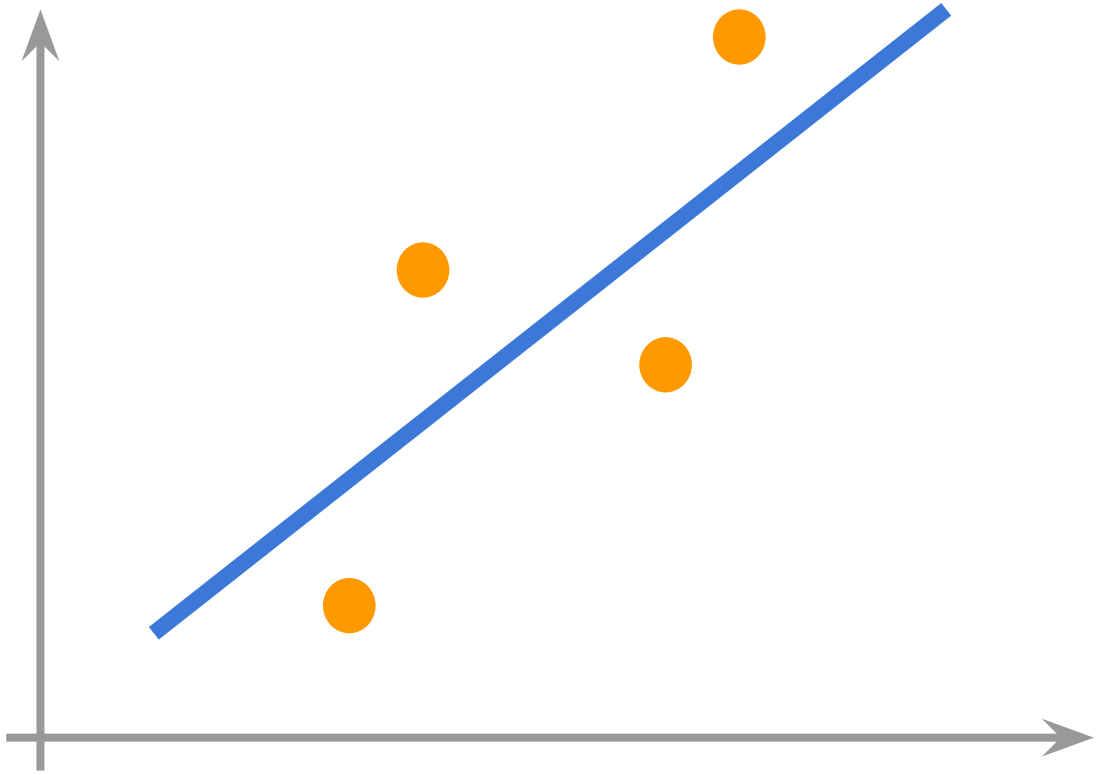
$$\textit{pred}\ y = 6\ (0) - 240 = -\ 240 \text{ pounds}$$

Least Squares Criterion

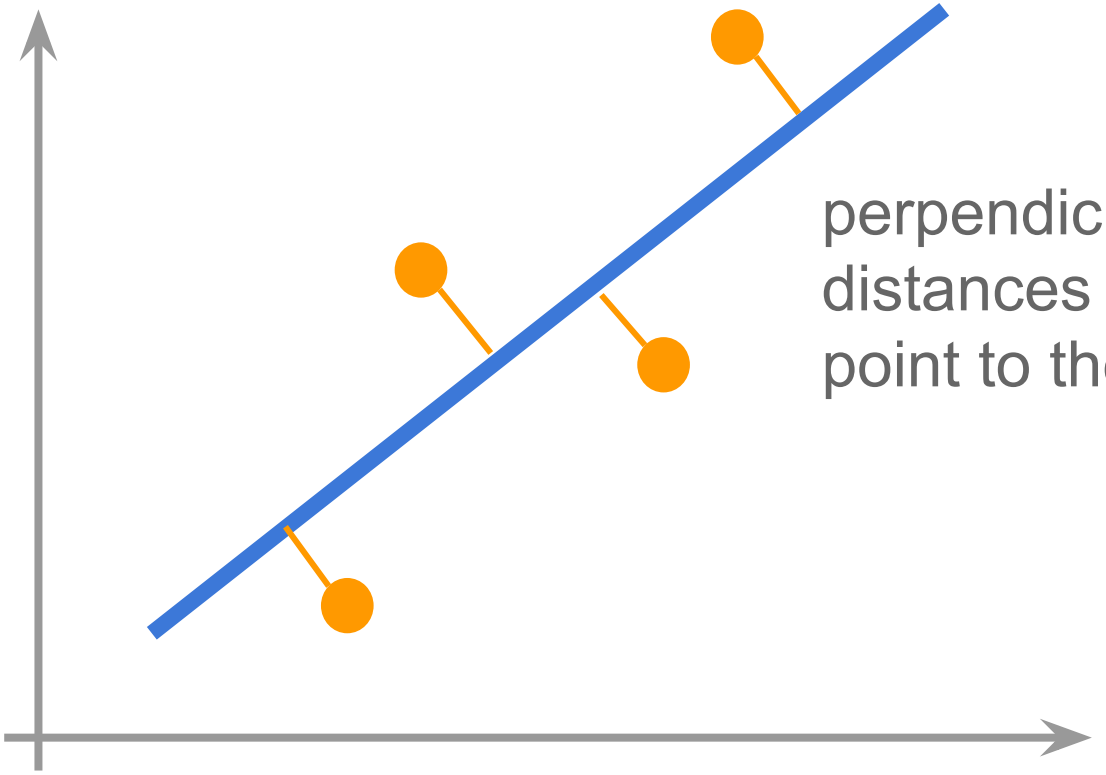
We need a mathematical criterion to find the "best" line

what do we mean by "best" line?

What criterion to fit a line?

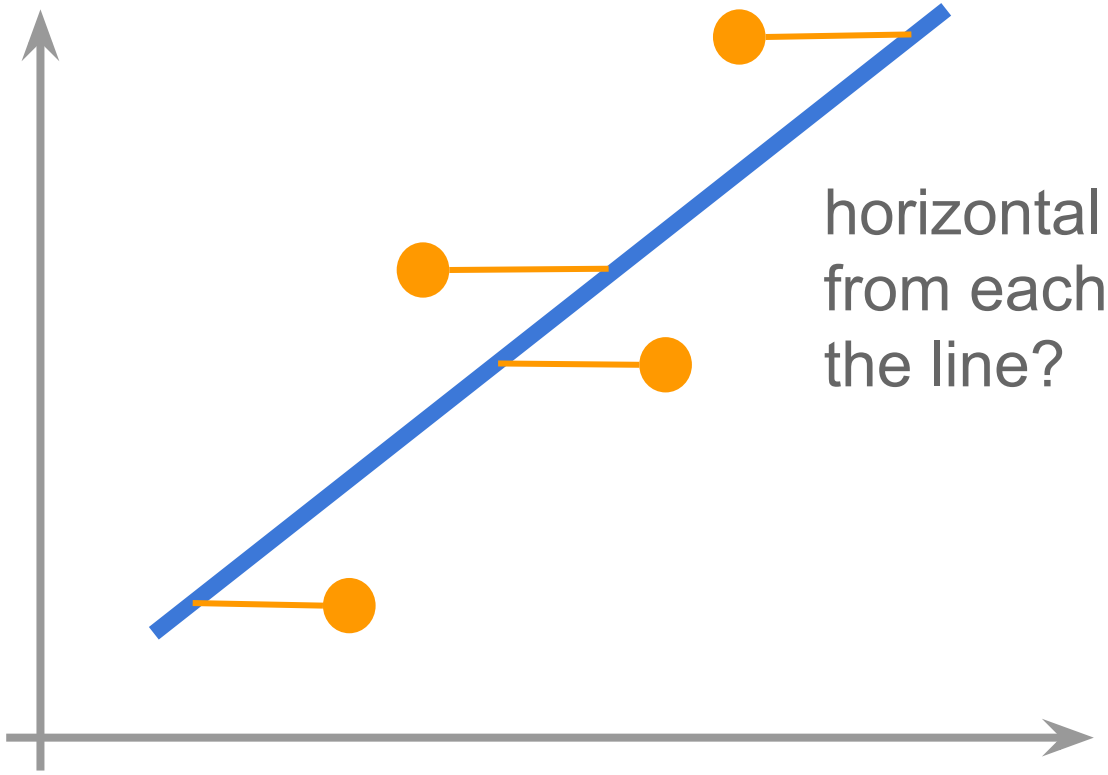


What criterion to fit a line?



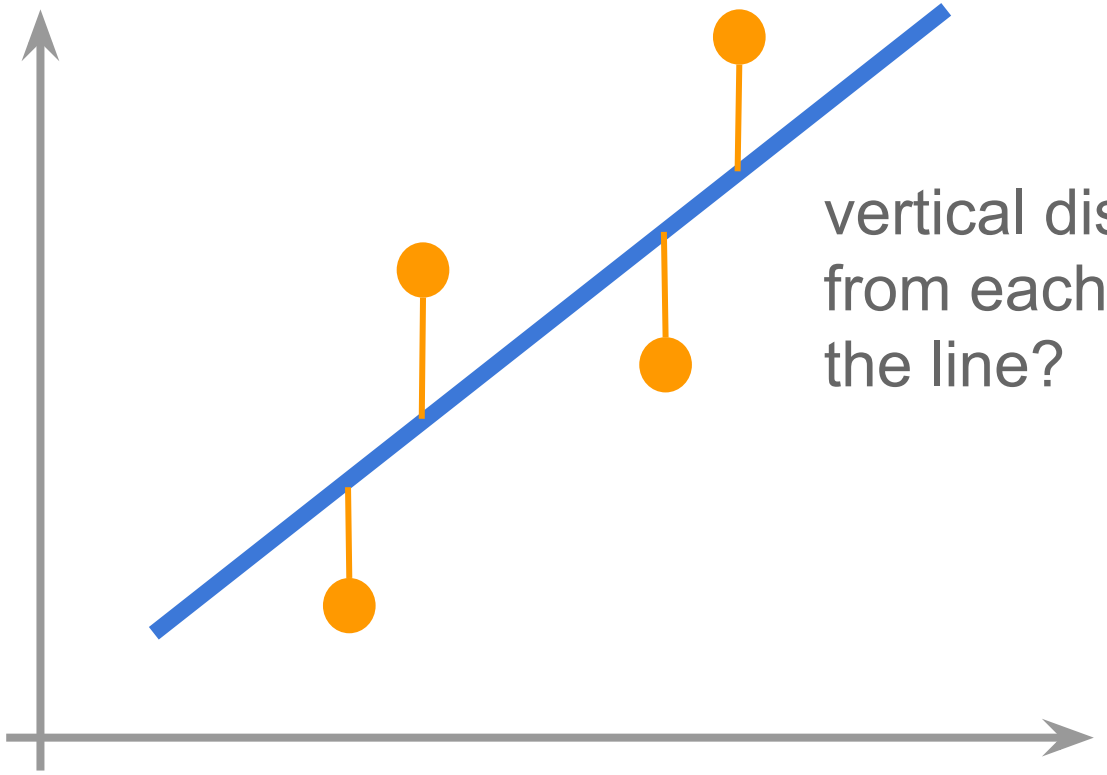
perpendicular distances from each point to the line?

What criterion to fit a line?



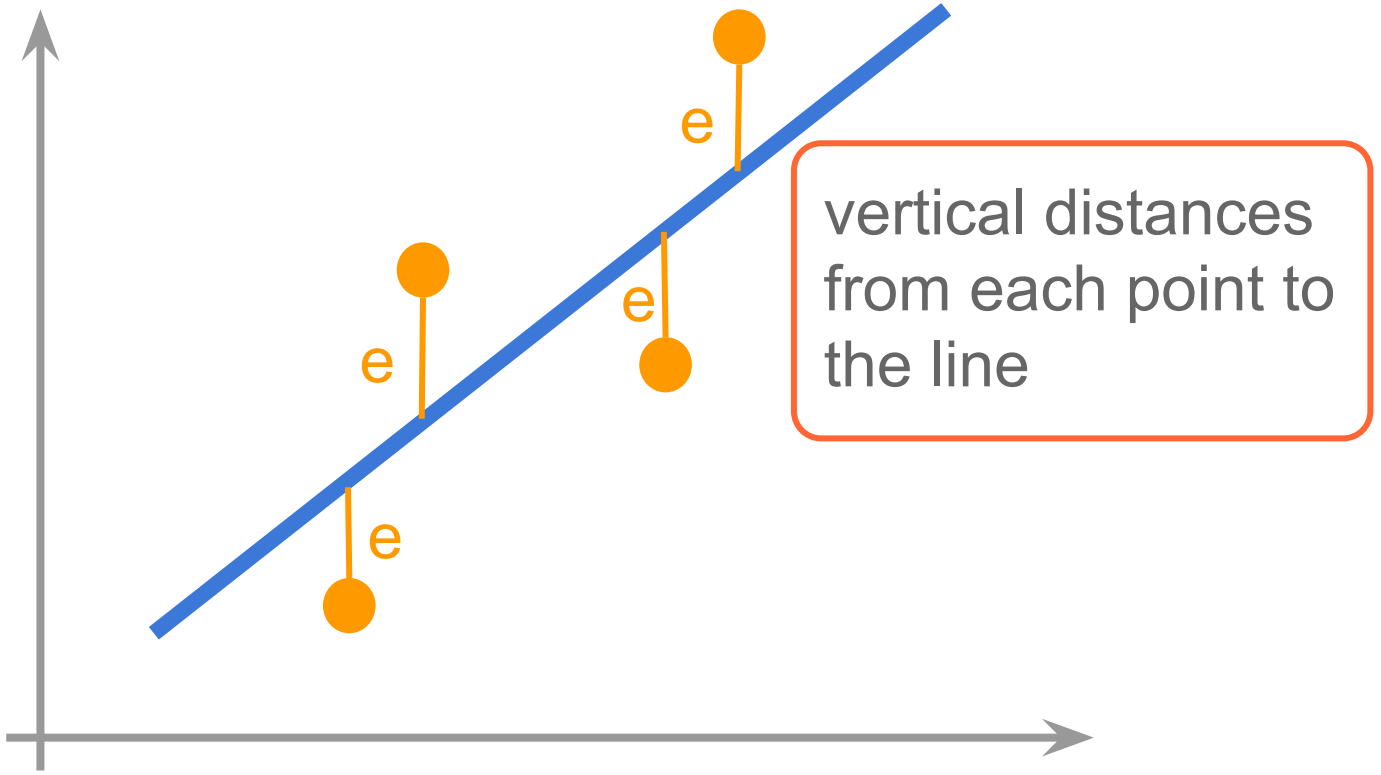
horizontal distances from each point to the line?

What criterion to fit a line?



vertical distances from each point to the line?

What criterion to fit a line?



What minimization criterion?

minimize Σe_i ?

sum of vertical distances?

What minimization criterion?

minimize Σe_i ?

sum of vertical distances?

minimize $\Sigma |e_i|$?

sum of absolute values of vertical distances?

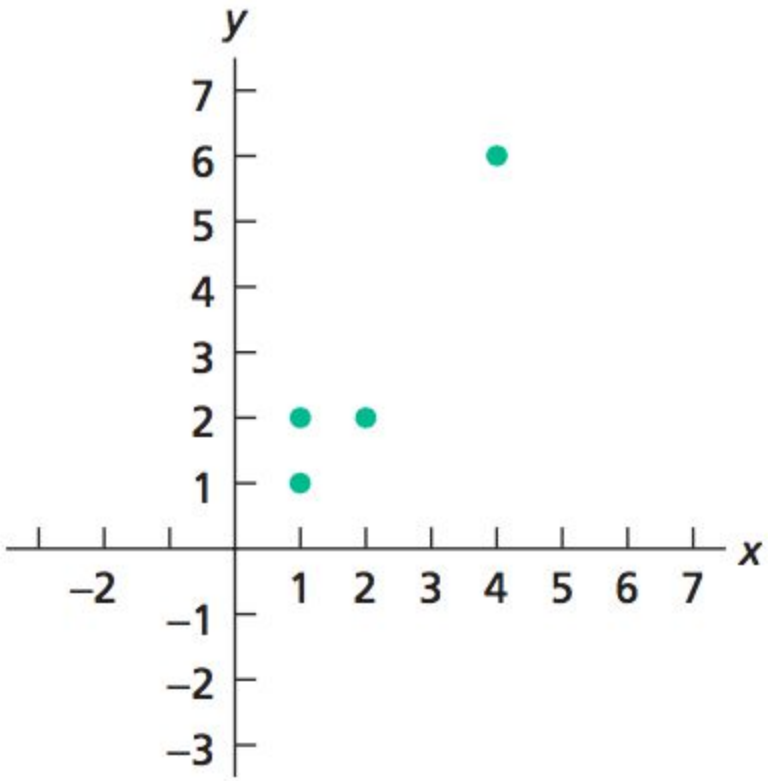
What minimization criterion?

minimize Σe_i ? sum of vertical distances?

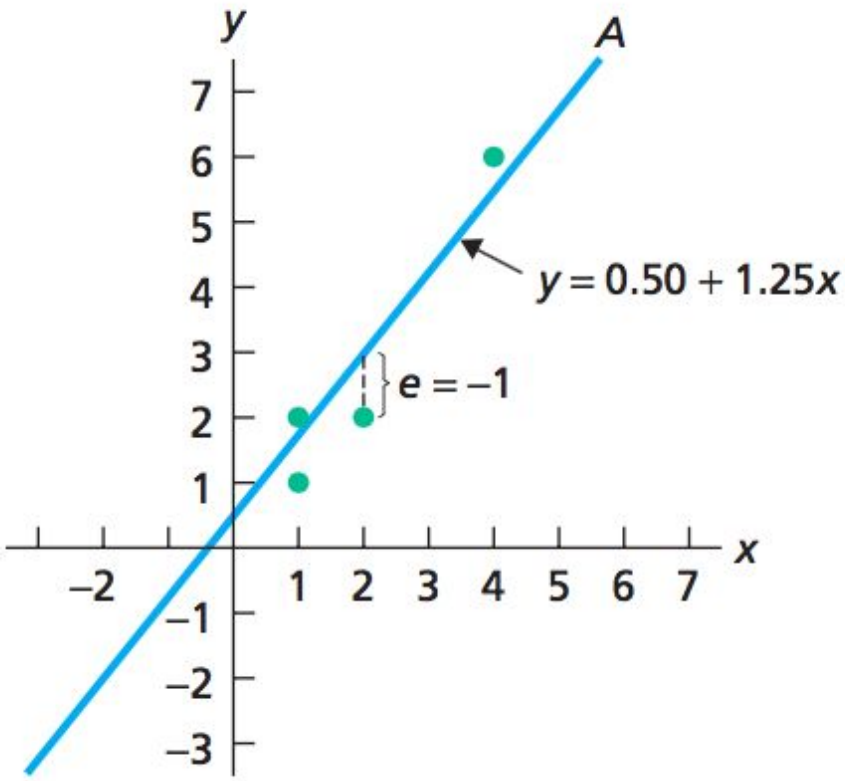
minimize $\Sigma |e_i|$? sum of absolute values of vertical distances?

minimize Σe_i^2 ? sum of squares of vertical distances?

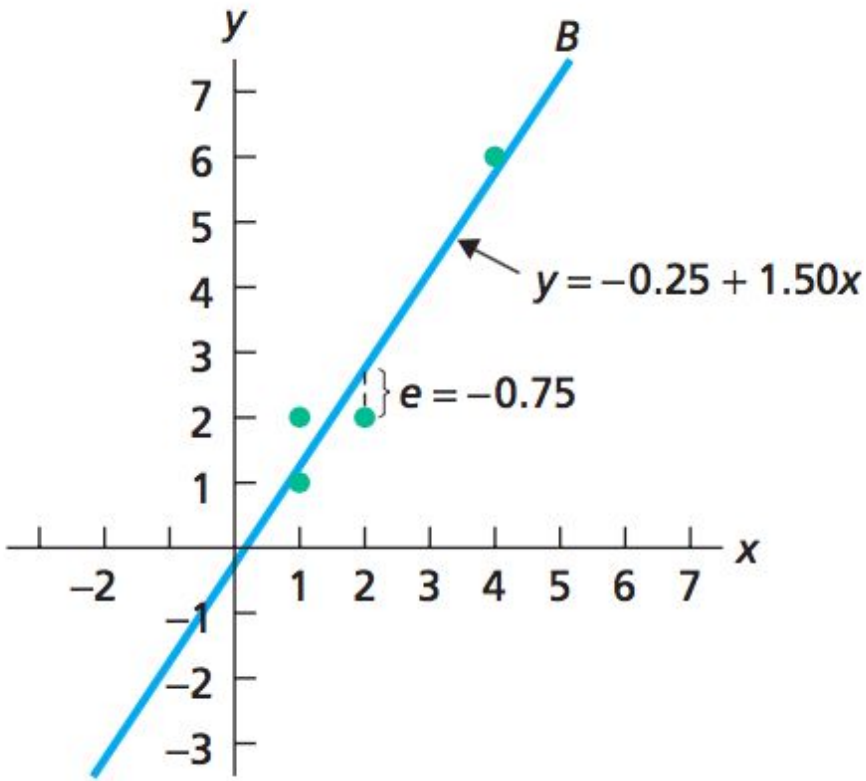
least squares criterion


y
7
6
5
4
3
2
1
−2
1
2
3
4
5
6
7
x
−1
−2
−3

Line A: $y = 0.50 + 1.25x$



Line B: $y = -0.25 + 1.50x$



What line fits the data points the best?

Line A: $y = 0.50 + 1.25x$

x	y	$\hat{y}$	e	e^2
1	1	1.75	−0.75	0.5625
1	2	1.75	0.25	0.0625
2	2	3.00	−1.00	1.0000
4	6	5.50	0.50	0.2500
				1.8750

(a)

Line B: $y = -0.25 + 1.50x$

x	y	$\hat{y}$	e	e^2
1	1	1.25	−0.25	0.0625
1	2	1.25	0.75	0.5625
2	2	2.75	−0.75	0.5625
4	6	5.75	0.25	0.0625
				1.2500

(b)

minimum sum of squared distances

Least Squares Criterion

The sum of the squares of the vertical distances from the points to the line is made as small as possible.

Alternative Formula

Alternative formula to find **b** and **m**

$$y = mx + b$$

slope intercept

Intercept and slope formulas

$$m = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$b = \bar{y} - b\bar{x}$$

Example

X = size of caribou population (100s)

Y = size of wolf population

Example

caribou	wolves			
x	y	x^2	y^2	xy
30	66	900	4356	1980
34	79	1156	6241	2686
27	70	729	4900	1890
25	60	625	3600	1500
17	48	289	2304	816
23	55	529	3025	1265
20	60	400	3600	1200
$\Sigma x = 176$	$\Sigma y = 438$	$\Sigma x^2 = 4628$	$\Sigma y^2 = 28{,}026$	$\Sigma xy = 11{,}337$

$\overline{x} = 25.14$ $\overline{y} = 62.57$

Slope

$$m = \frac{n\Sigma xy - (\Sigma x)\,(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$m = \frac{7(11{,}337) - (176)\,(438)}{7(4628) - (176)^2} = 1.60$$

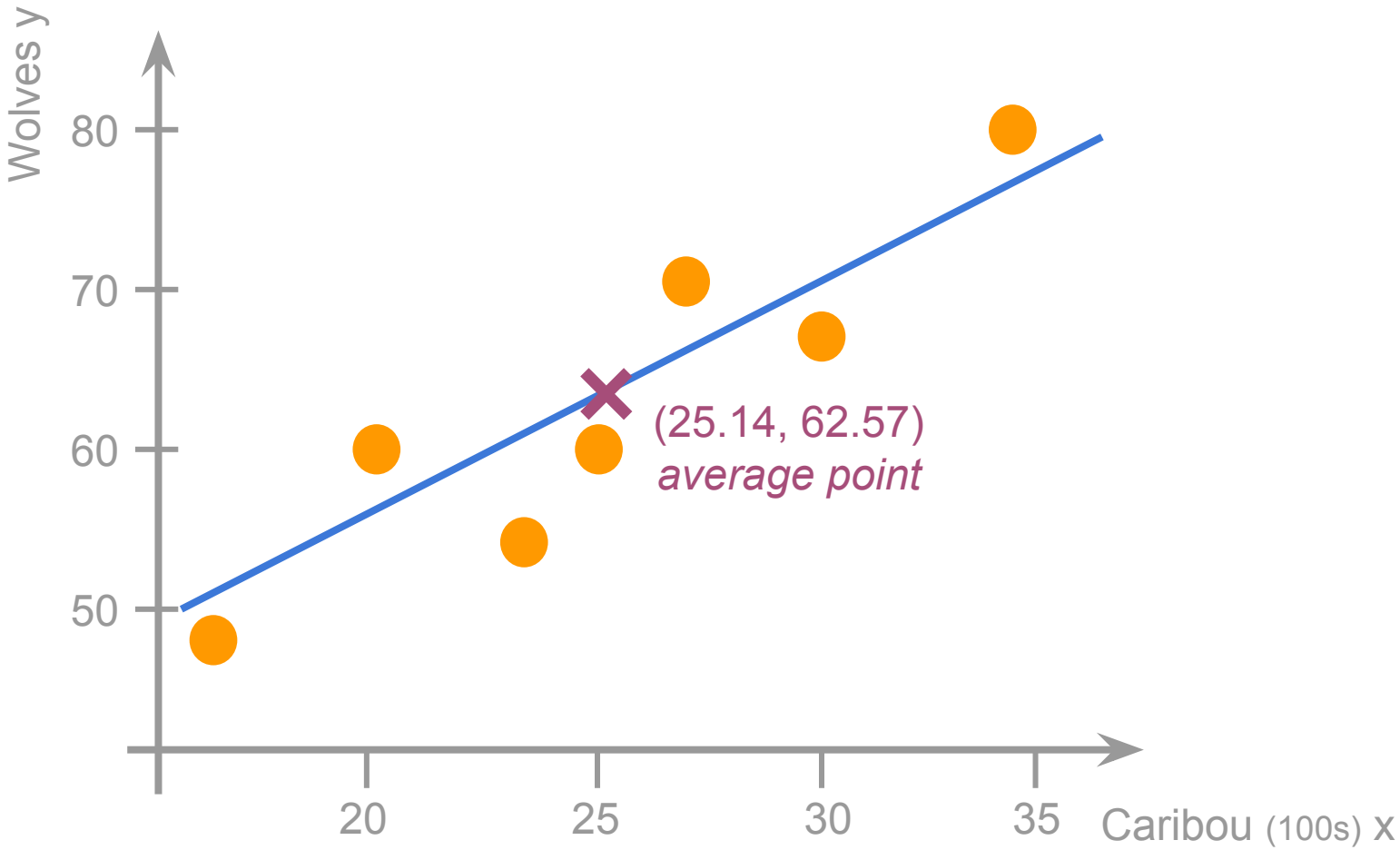
Intercept

$$b = \bar{y} - m\bar{x}$$

$$b = 62.57 - 1.60\ (25.14) = 22.35$$

$$\hat{y} = mx + b = 1.60\ x + 22.35$$

Regression line



Regression for prediction

We can use a regression model to make predictions

Prediction: Obtaining a value of **y** for a "new" value of **x**

Predicting y values

caribou	wolves
x	y
30	66
34	79
27	70
25	60
17	48
23	55
20	60
21	?

Caribou & Wolves example

For x = 21, what is the predicted $\hat{y}$?

$$\hat{y} = 22.35 + 1.60\ x$$

$$\hat{y} = 22.35 + 1.60\ (21) = 55.95$$

Predicted values

caribou	wolves
x	y
30	66
34	79
27	70
25	60
17	48
23	55
20	60
21	56