Introduction to Regression

Gaston Sanchez

Motivation

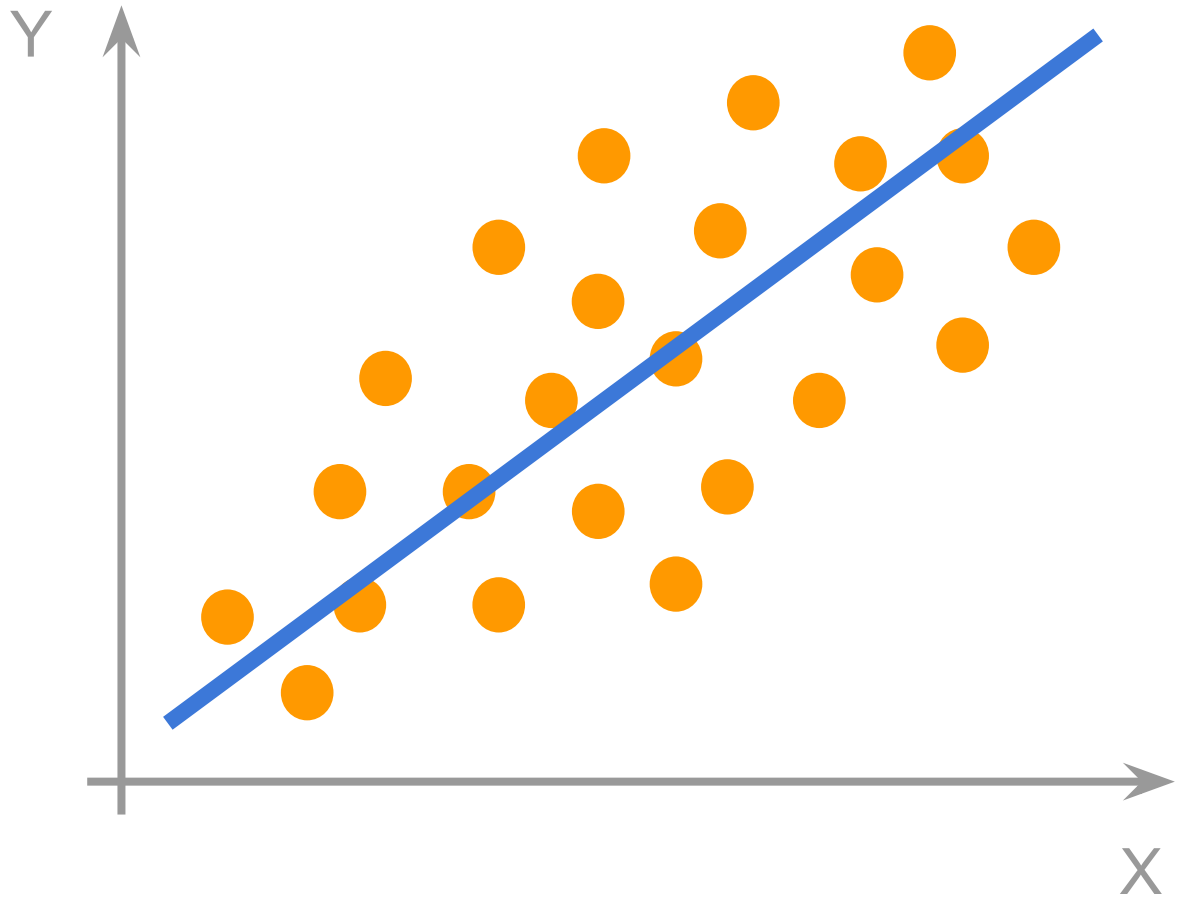
Regression Idea

Summarize the relationship between X and Y with a linear equation

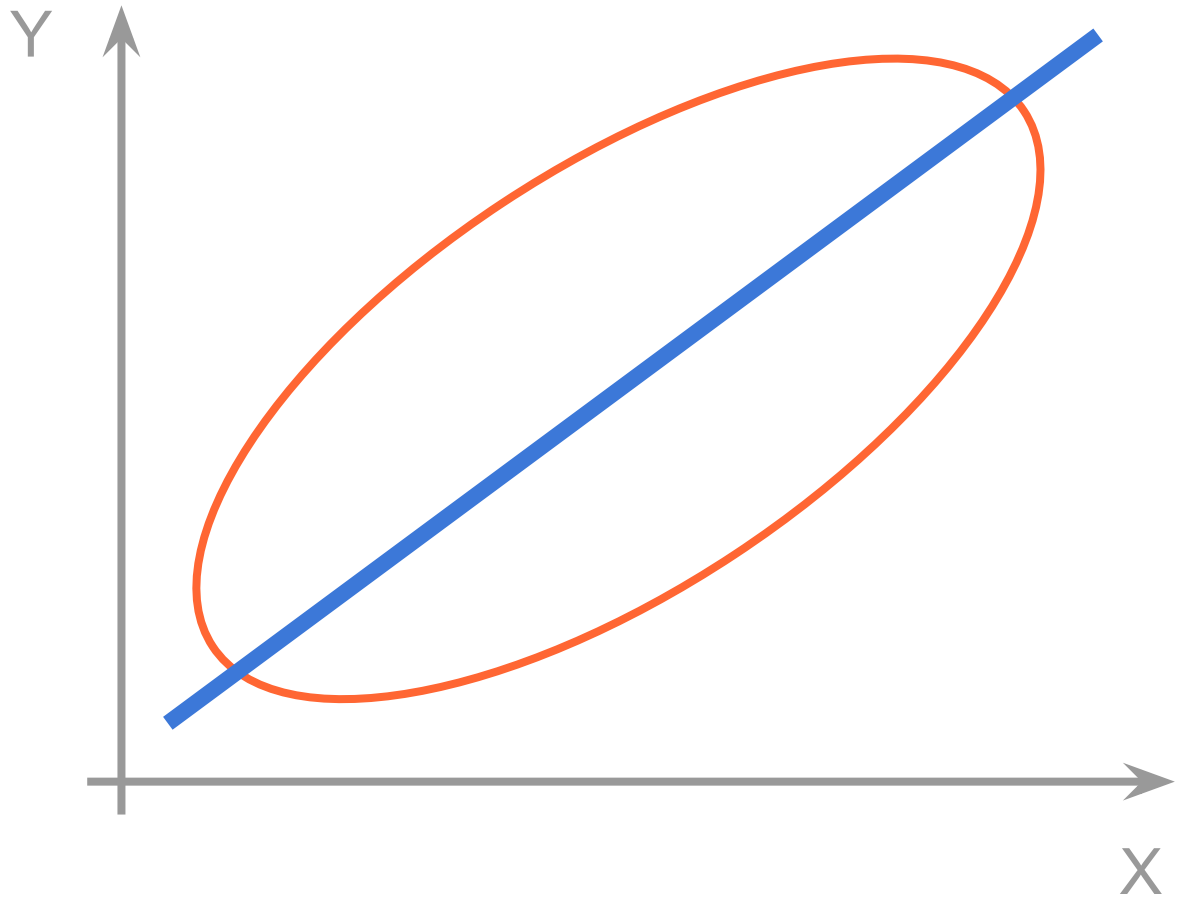
In graphical terms ...

Represent a cloud of points with a line

Cloud of points



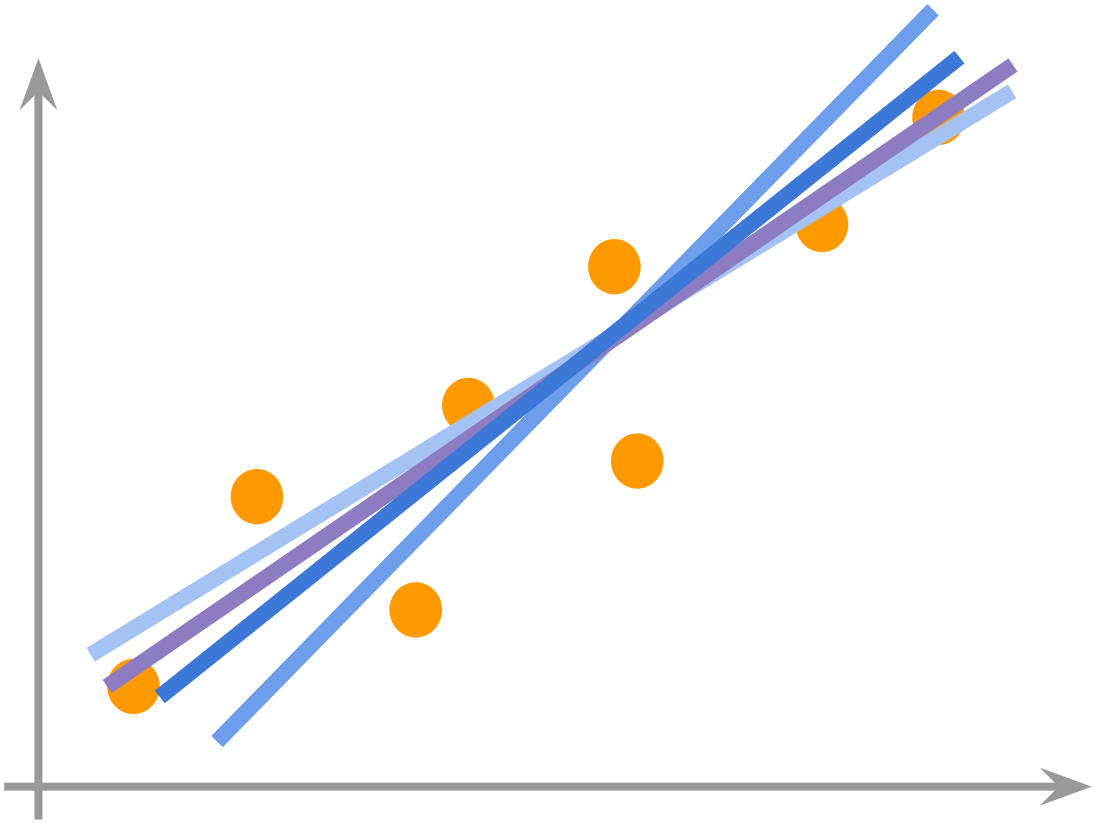
Summarize cloud with a line



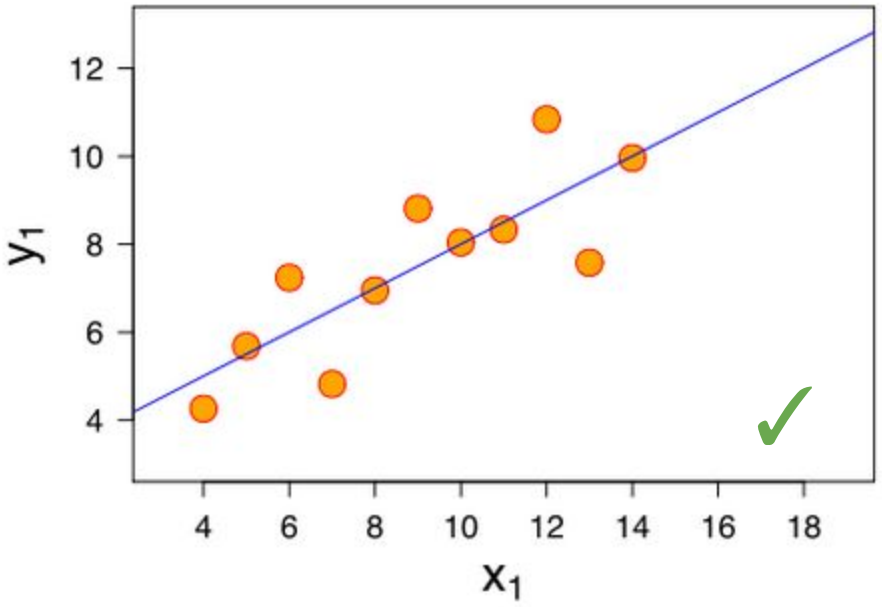
How to find the "best" fitting line?

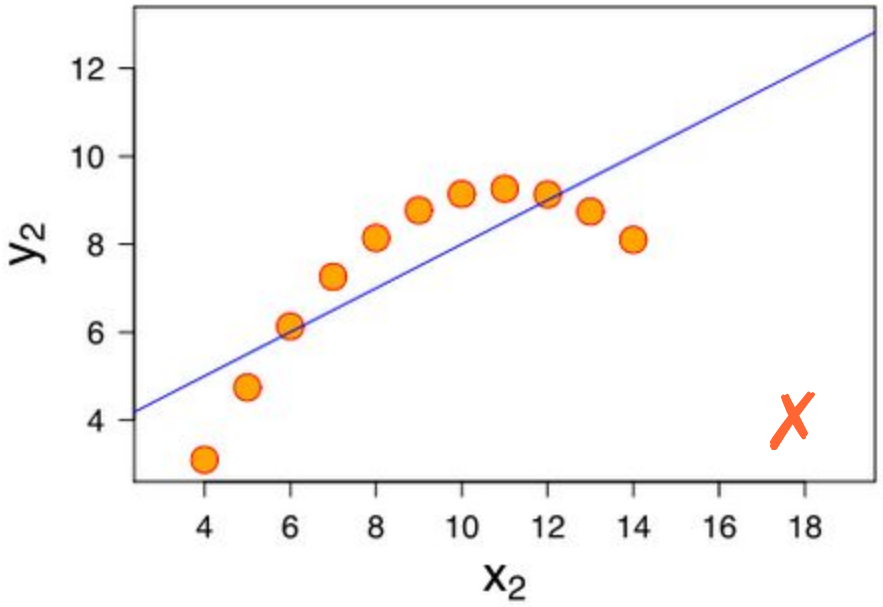
Fitting a line

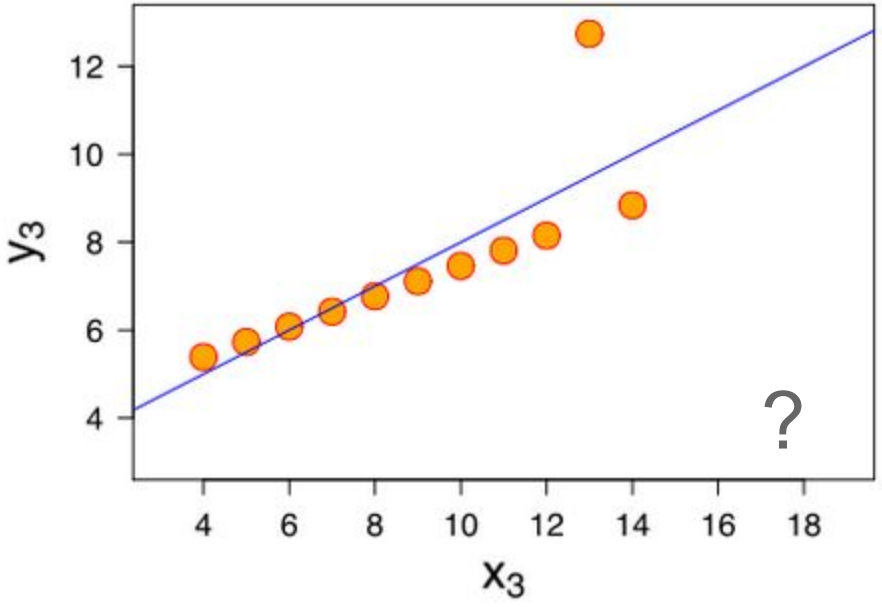
which is the best line?

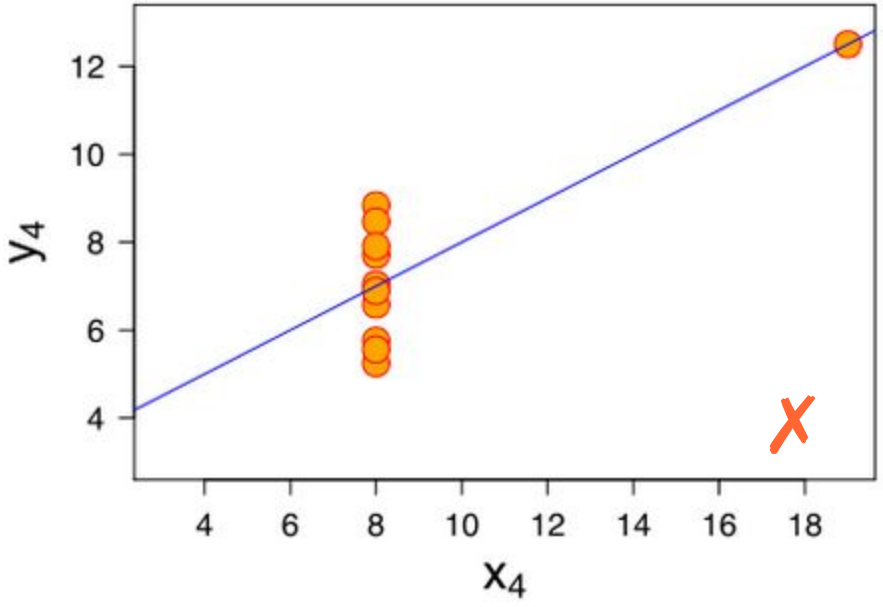


Warning on the use of a regression line

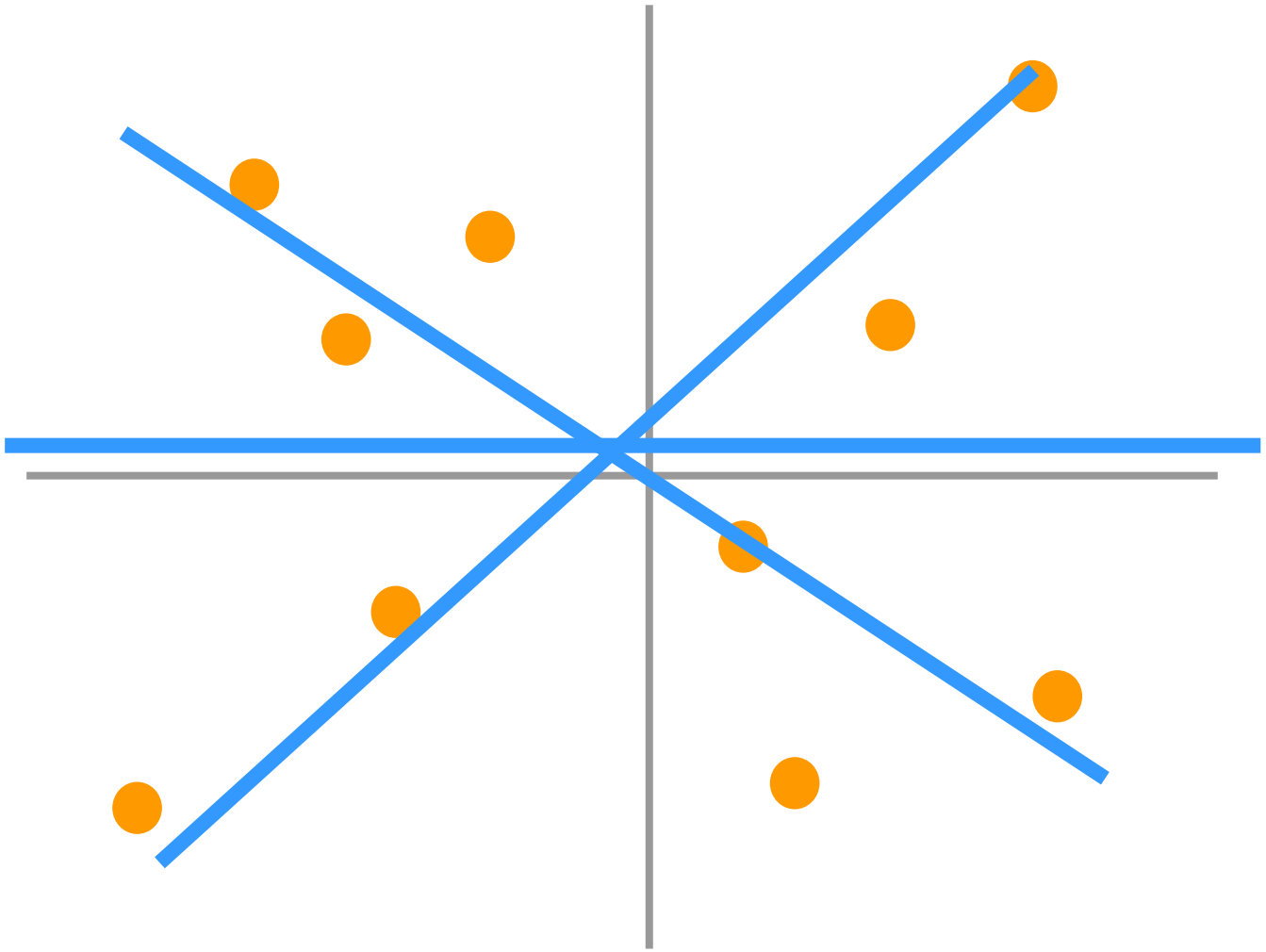

y1
x1
✓


y2
x2
✗


y3
x3
?


y4
x4
✗

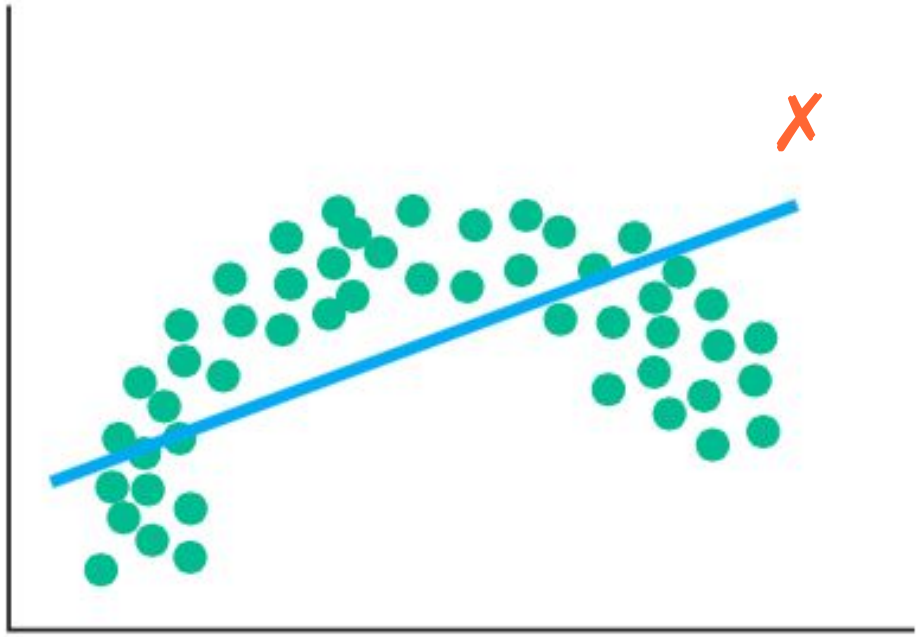
Which line?



Warning

The idea behind a regression line is based on the assumption that the data points are scattered about a line.

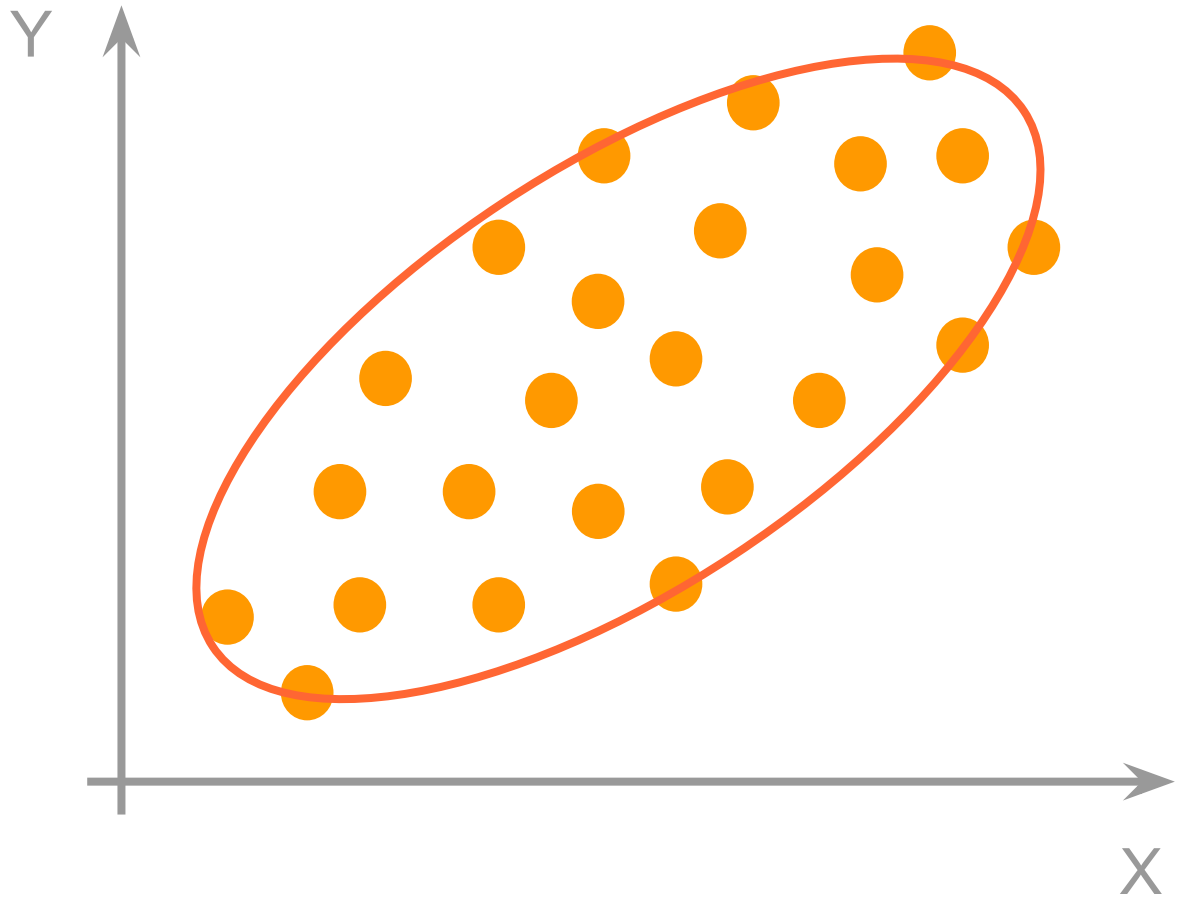
Warning



Demo: Heights of parents and children

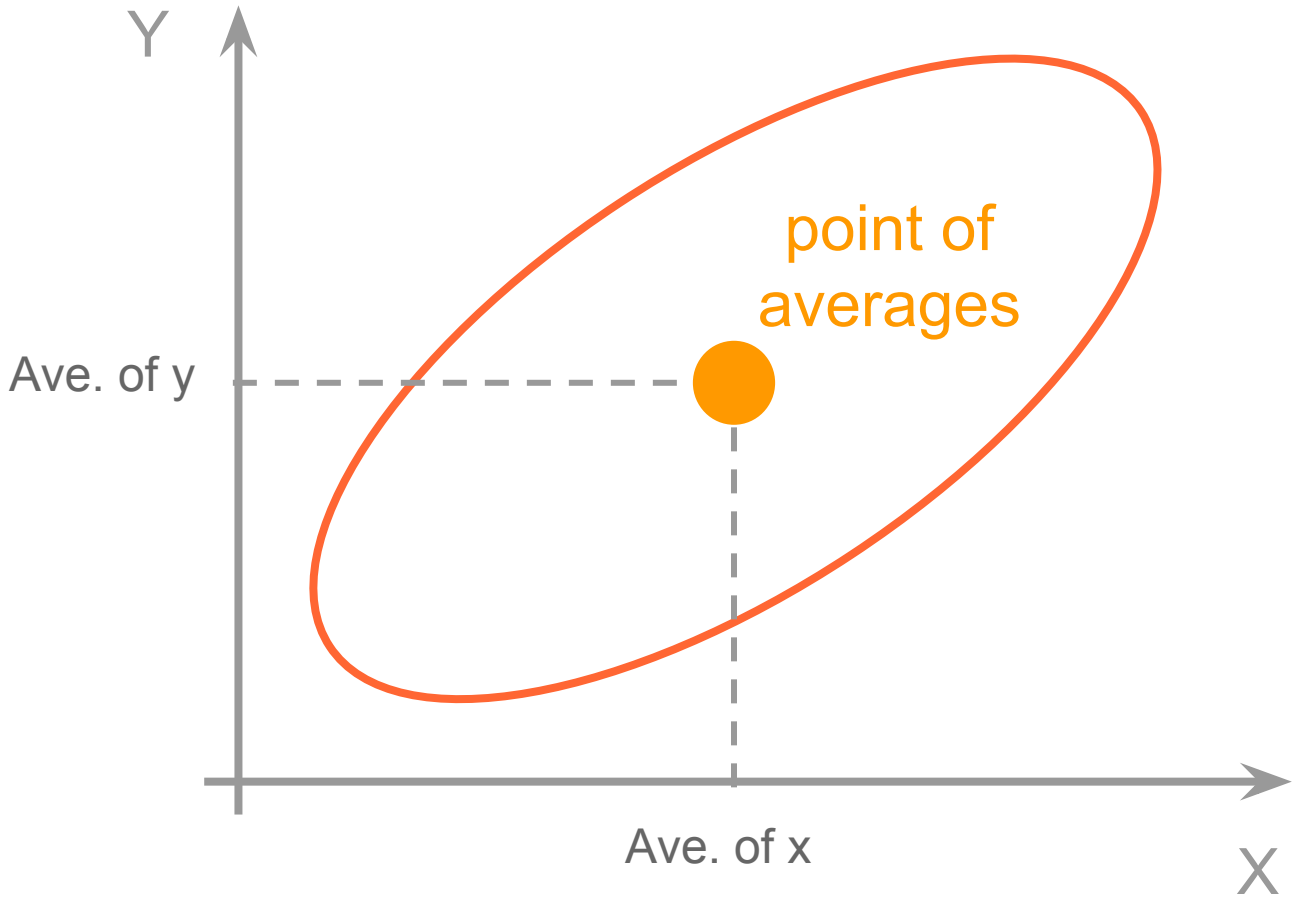
Cloud of Points

Cloud of points

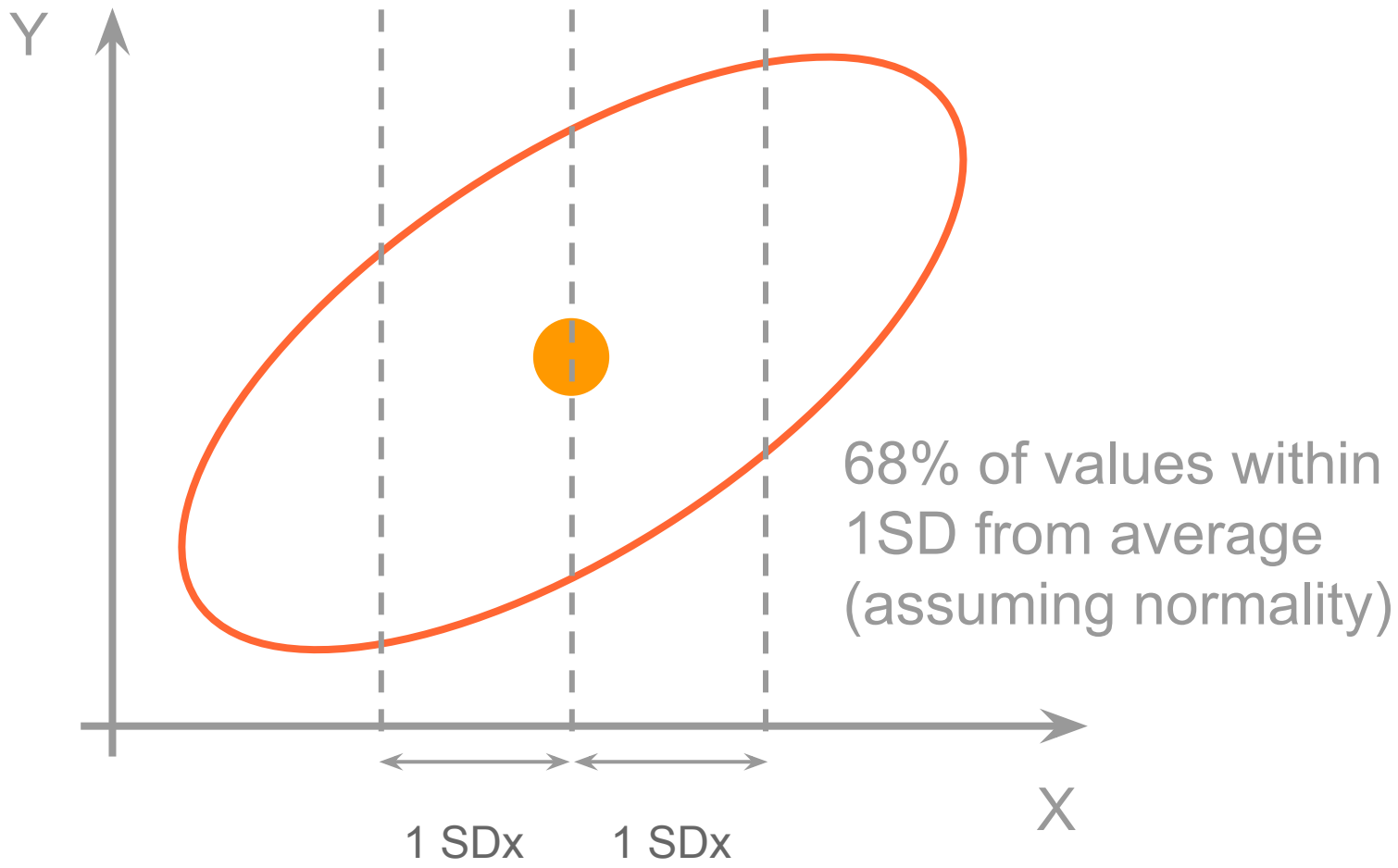


Overall shape

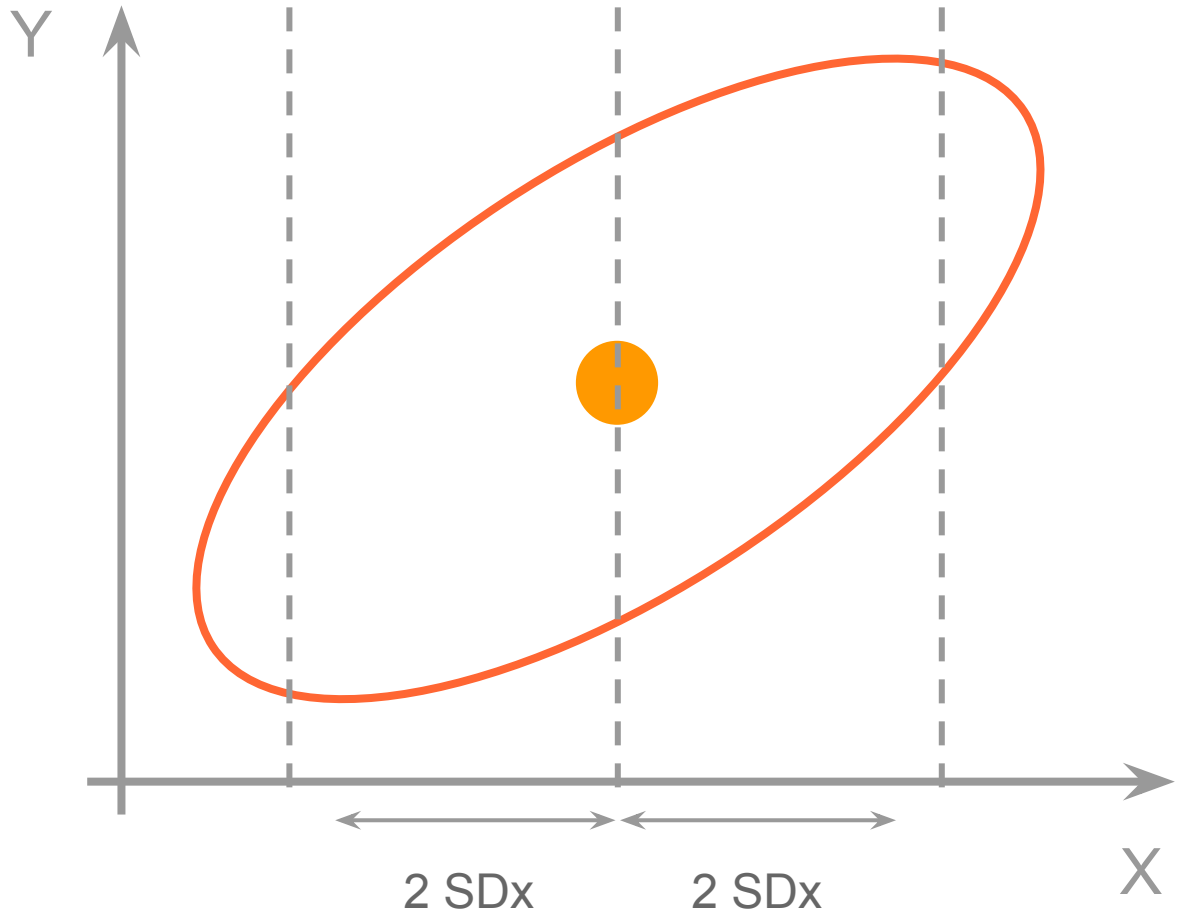
The point of averages



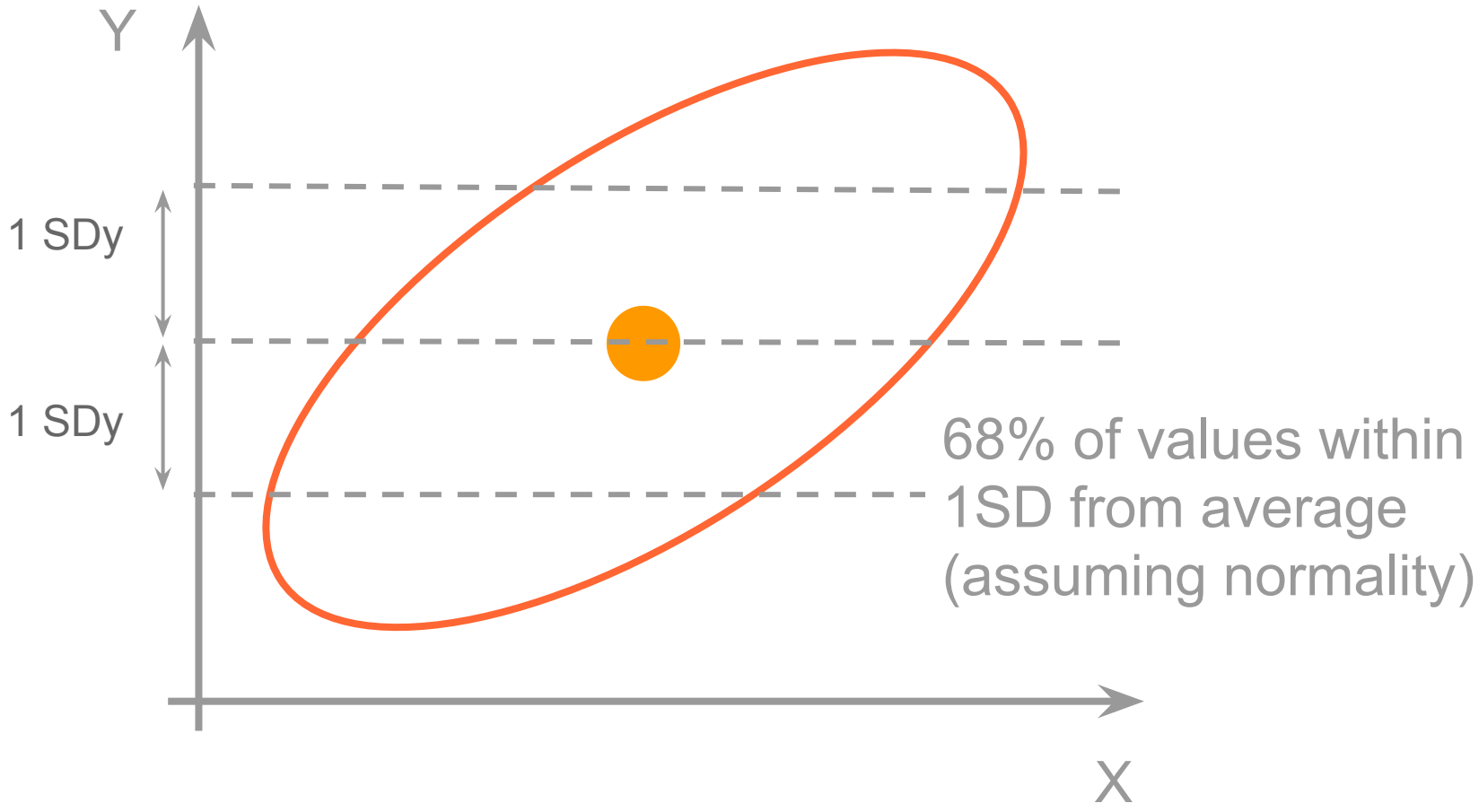
The horizontal SD



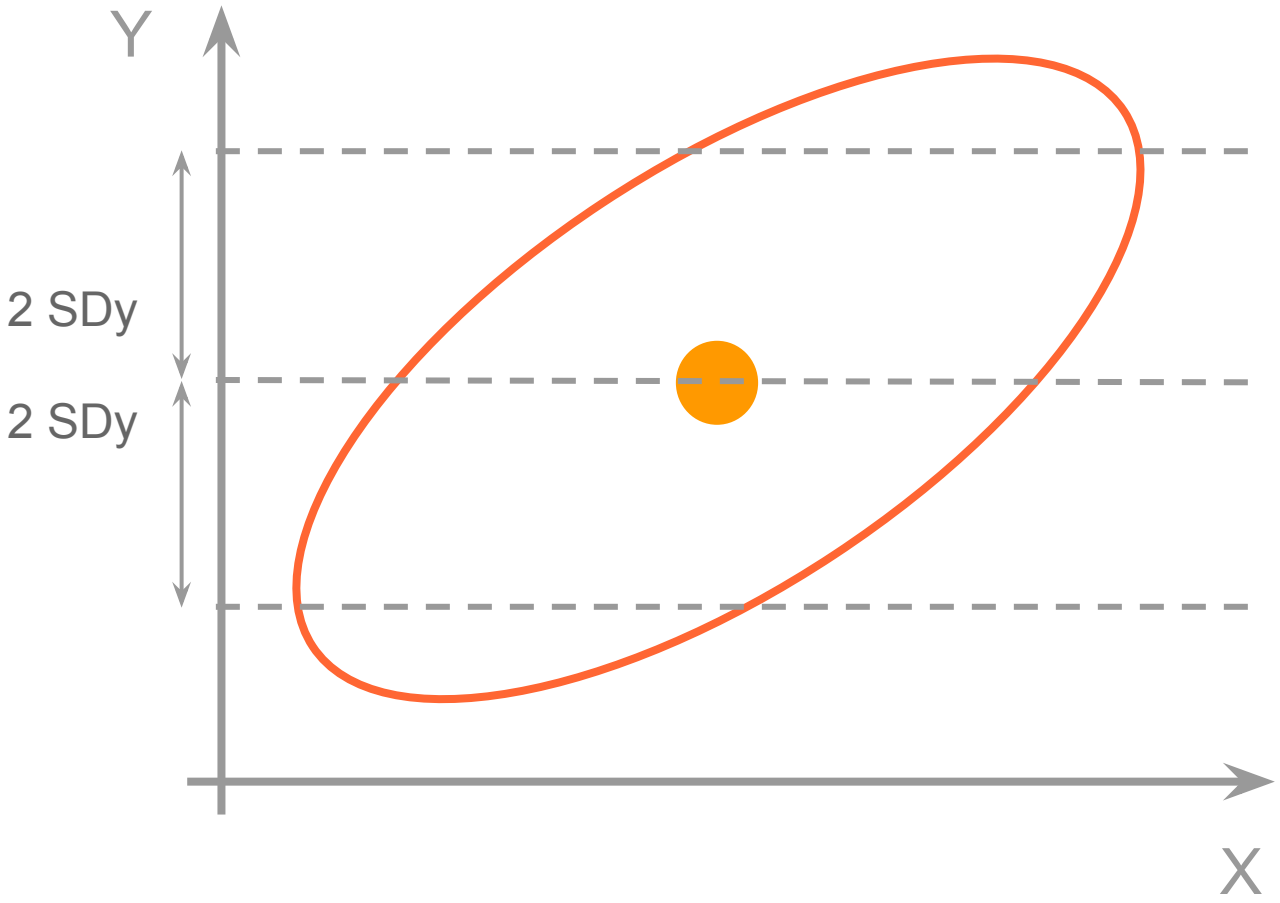
The horizontal SD



The vertical SD



The vertical SD



About the SD line:

Passes through the point of averages

If **r** > 0, the SD line goes up by 1 SDy when X goes 1 SDx to the right

Systematically off from the graph of averages

SD line is NOT the Regression line

Not really a good summary of the data

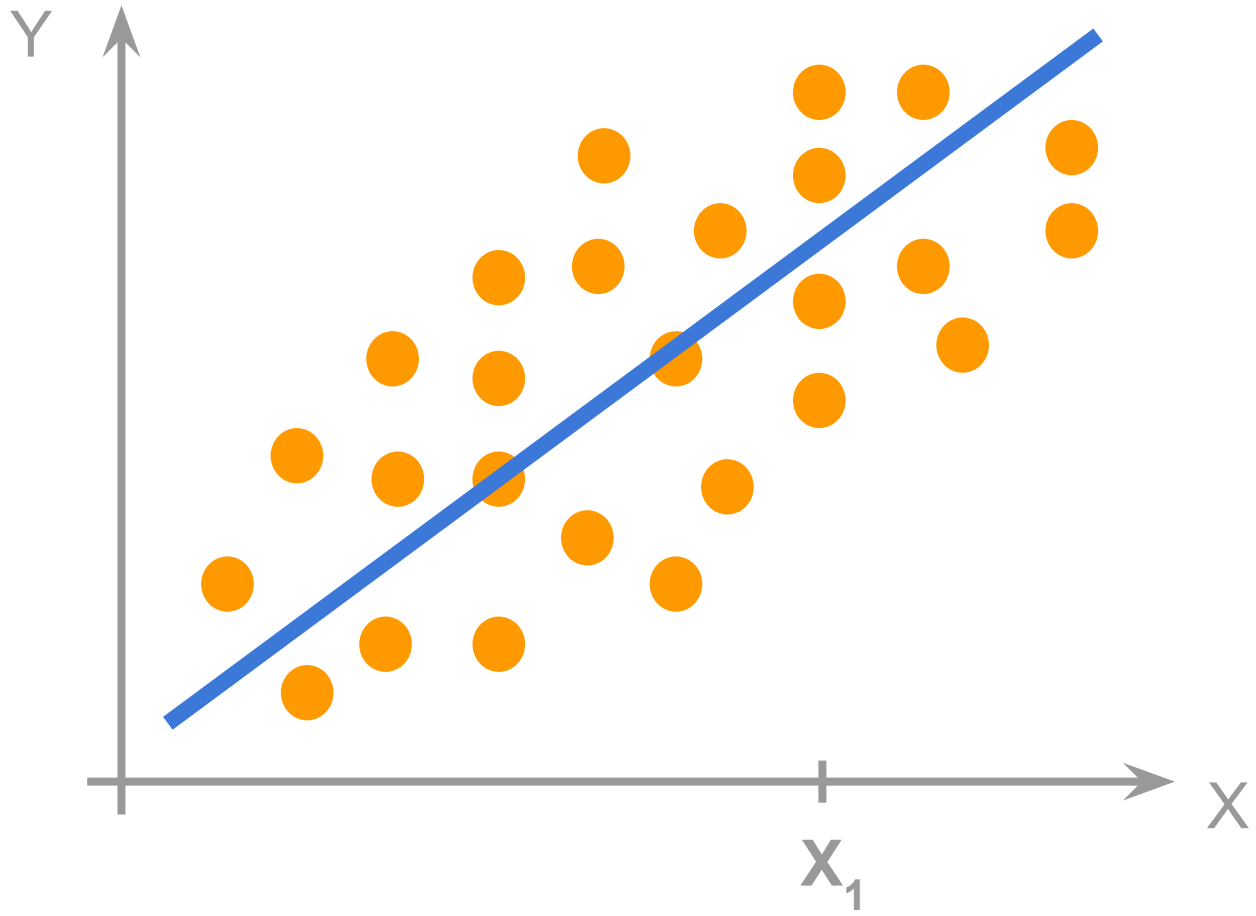
Desirable Condition

A “good” summary line would pass very close to the graph of averages

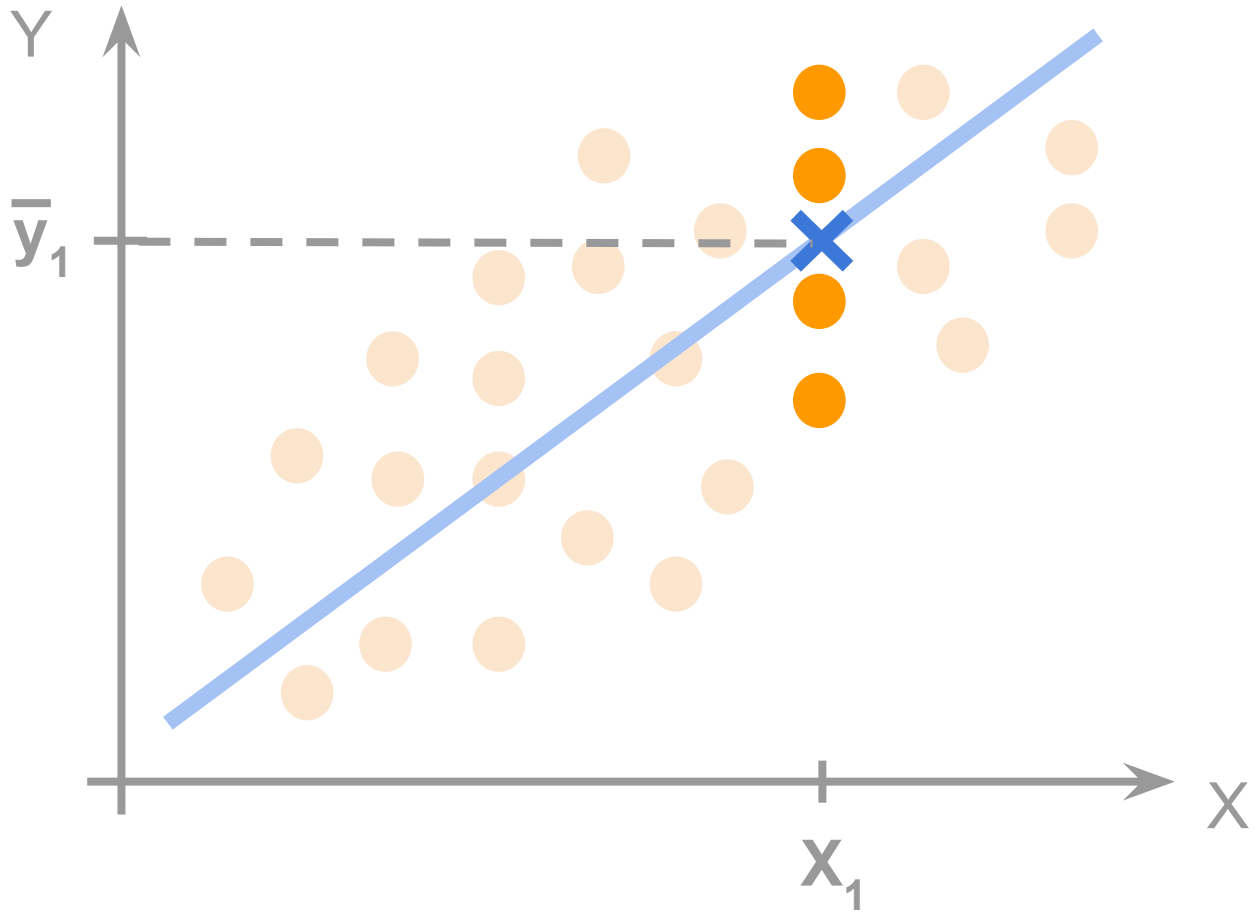
Regression Method

The regression line for **y** on **x** estimates the average value for **y** corresponding to each value of **x**

X estimates average of Y

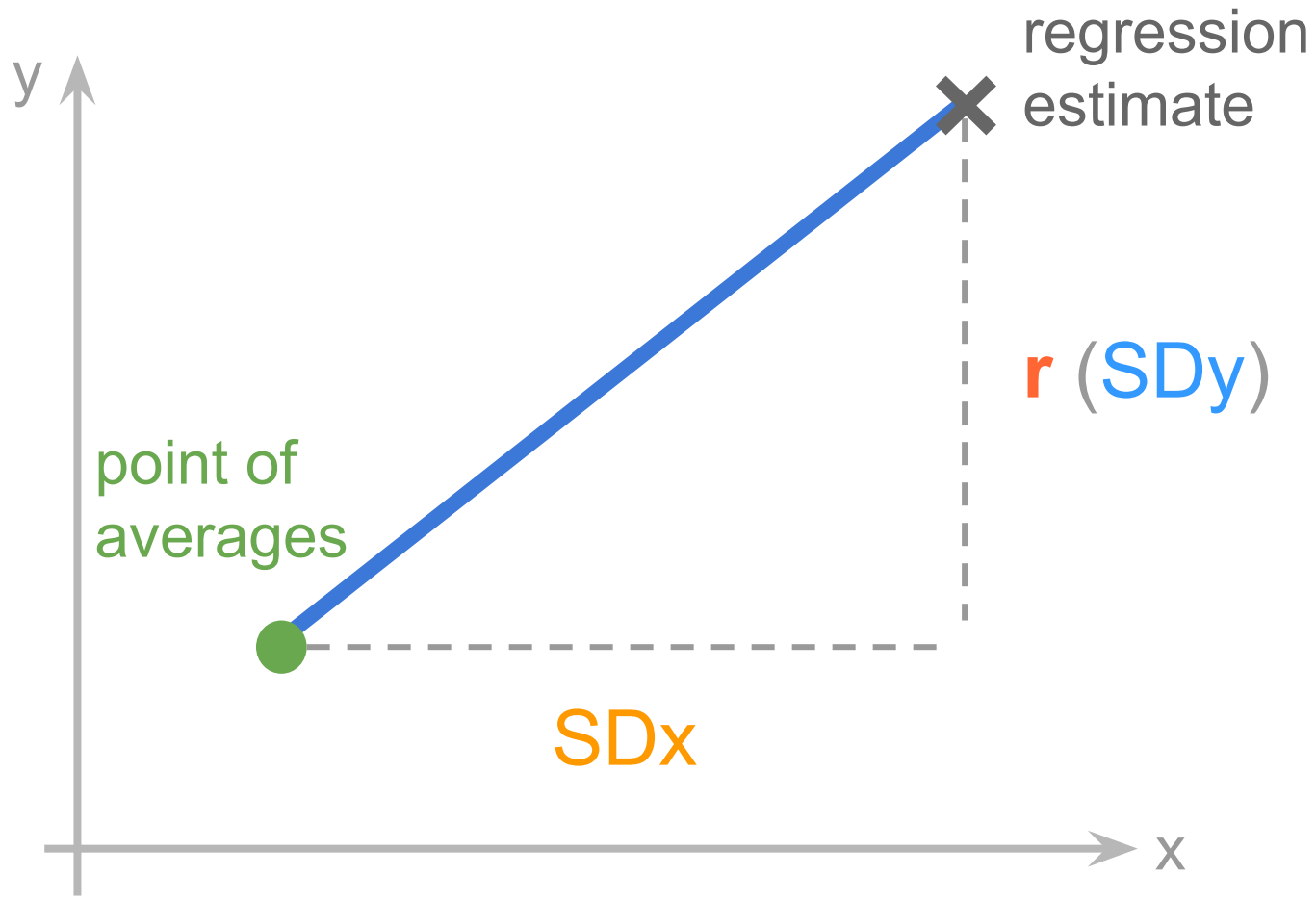


X estimates average of Y



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 Method for Individuals

Group of men with:

- average height = **70** inches, SD of **3** inches
- average weight = **180** pounds, SD of **45** pounds
- r = 0.40

Using the regression method to predict the weight of a man

Regression Method for Individuals

Group of men with:

- X: average height = **70** in, SD of **3** in
- Y: average weight = **180** lbs, SD of **45** lbs
- r = 0.40

Weight of man
(unknown height)

best guess:
180 pounds

Predicting weight from height

Group of men with:

- X: average height = **70** in, SD of **3** in
- Y: average weight = **180** lbs, SD of **45** lbs
- r = 0.40

Weight of man
with height = 73 in

73 in is 1 SDx
above avg 70 in

Predicting weight from height

73 inches is 3 inches above the average (70)

i.e. 1 SDx above the average

$$\text{weight} = (\mathbf{180}) + (\mathbf{0.4})\,(\mathbf{45}) = 198$$

r SDy

The Regression Effect

Regression effect:

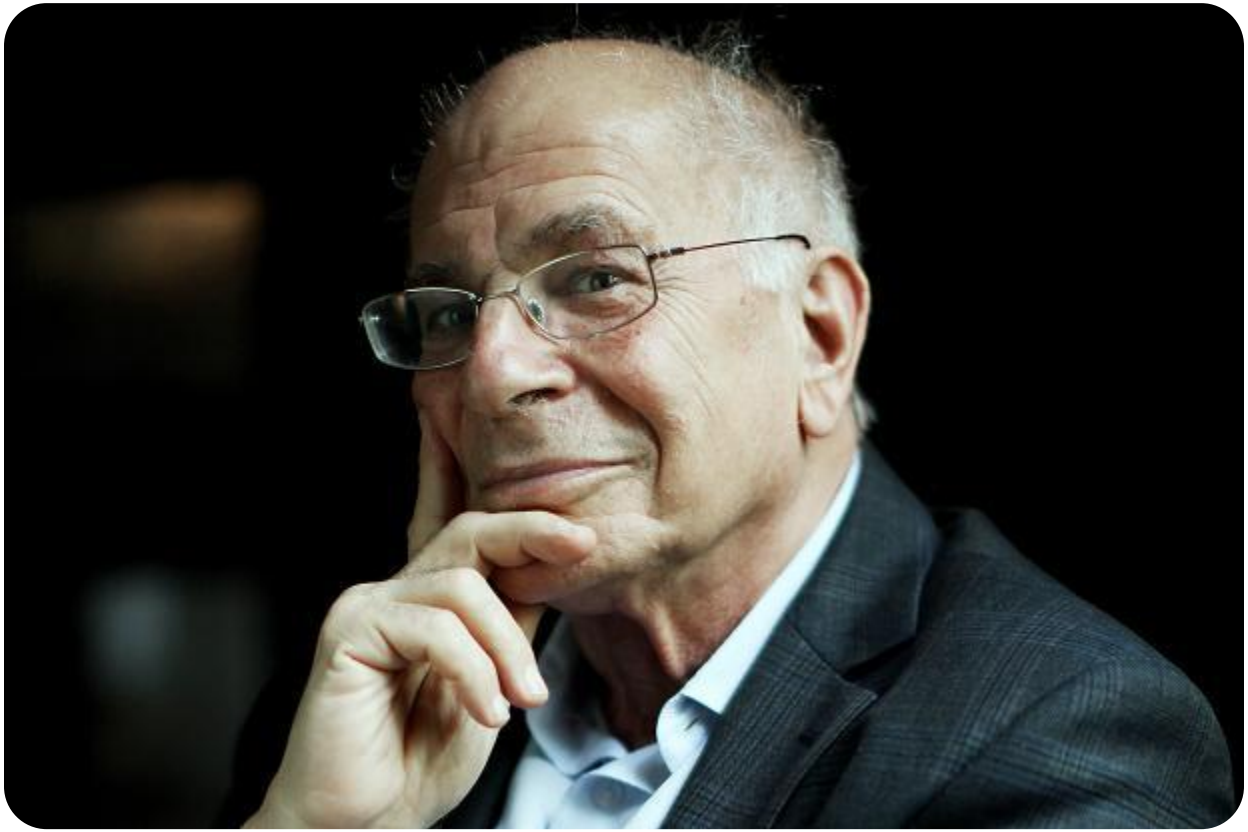
Events tend to regress towards average over time

The Regression Effect

In virtually all *test-retest* situations, the bottom group on the first test will on average show some improvement on the second test, and the top group will on average fall back.

Demo: NBA Point Guards

Daniel Kahneman



Kahneman's example

Israeli-American psychologist

2002 Nobel Prize in Economic Sciences

Example: Psychology department of the Israeli Air Force

Flight instructors praise -vs- blame

Instructor’s comment:

"On many occasions I have praised flight cadets for clean execution of some aerobatic maneuver, and in general when they try it again, they do worse. On the other hand, I have often screamed at cadets for bad execution, and in general they do better the next time. So please don't tell us that reinforcement works and punishment does not, because the opposite is the case."