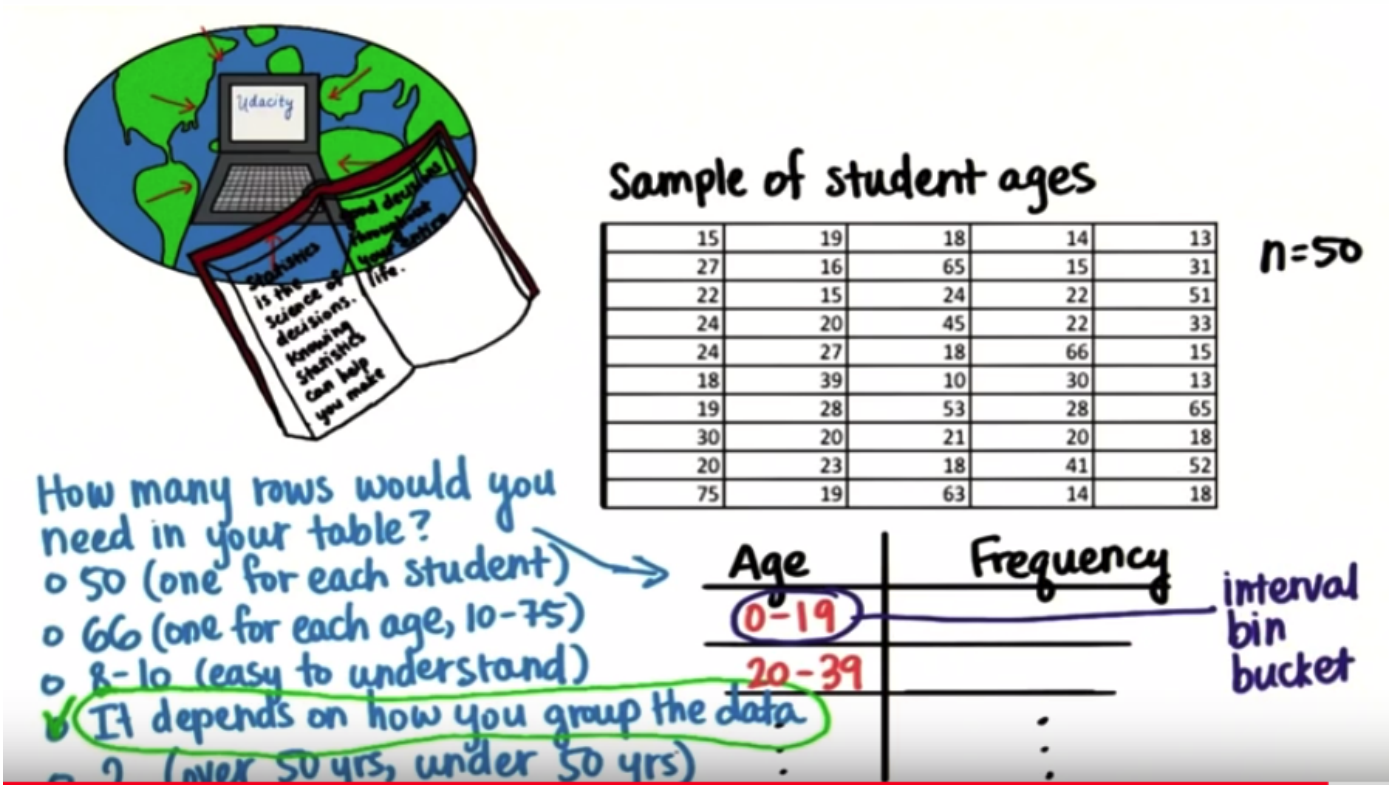


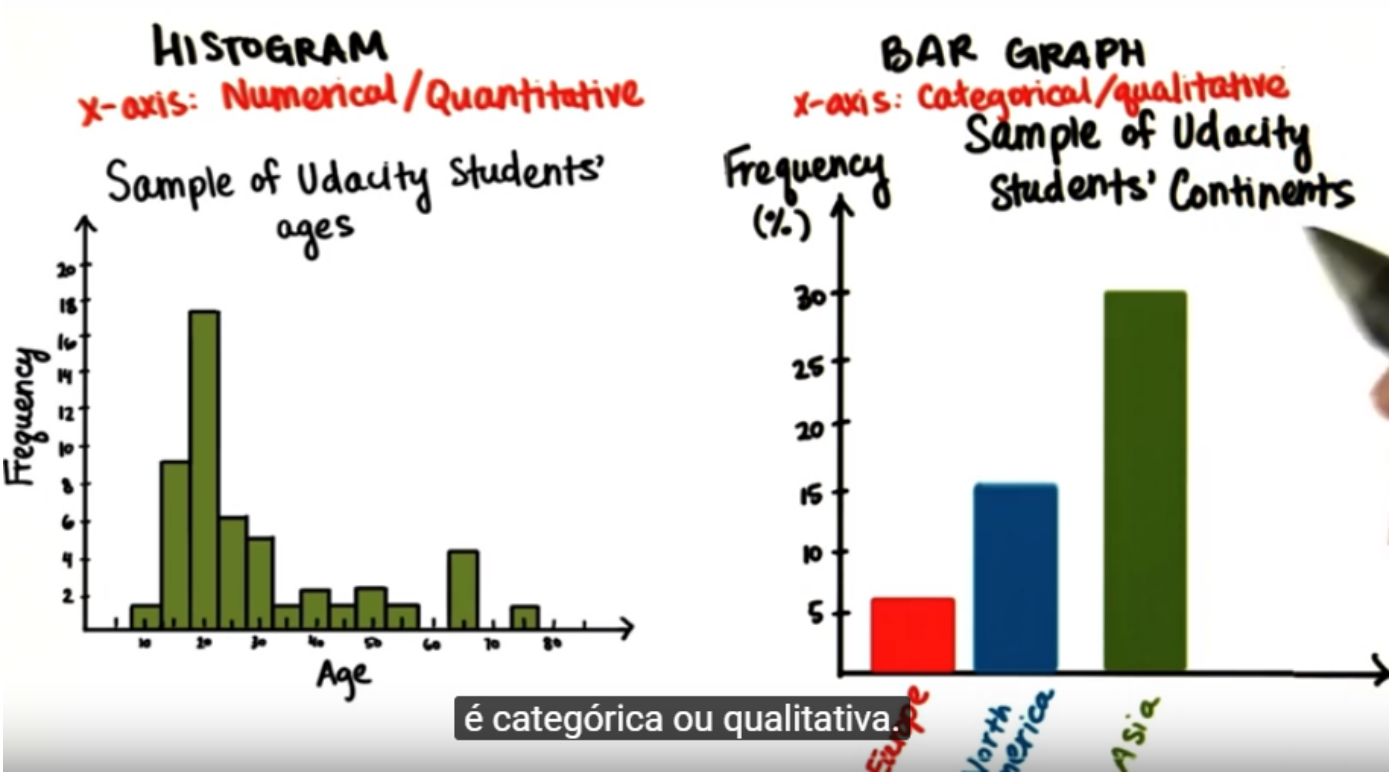
Sobre os eixos do grafico, a frequencia/variável dependente/resultado sempre fica no axes Y e a variável no axes X

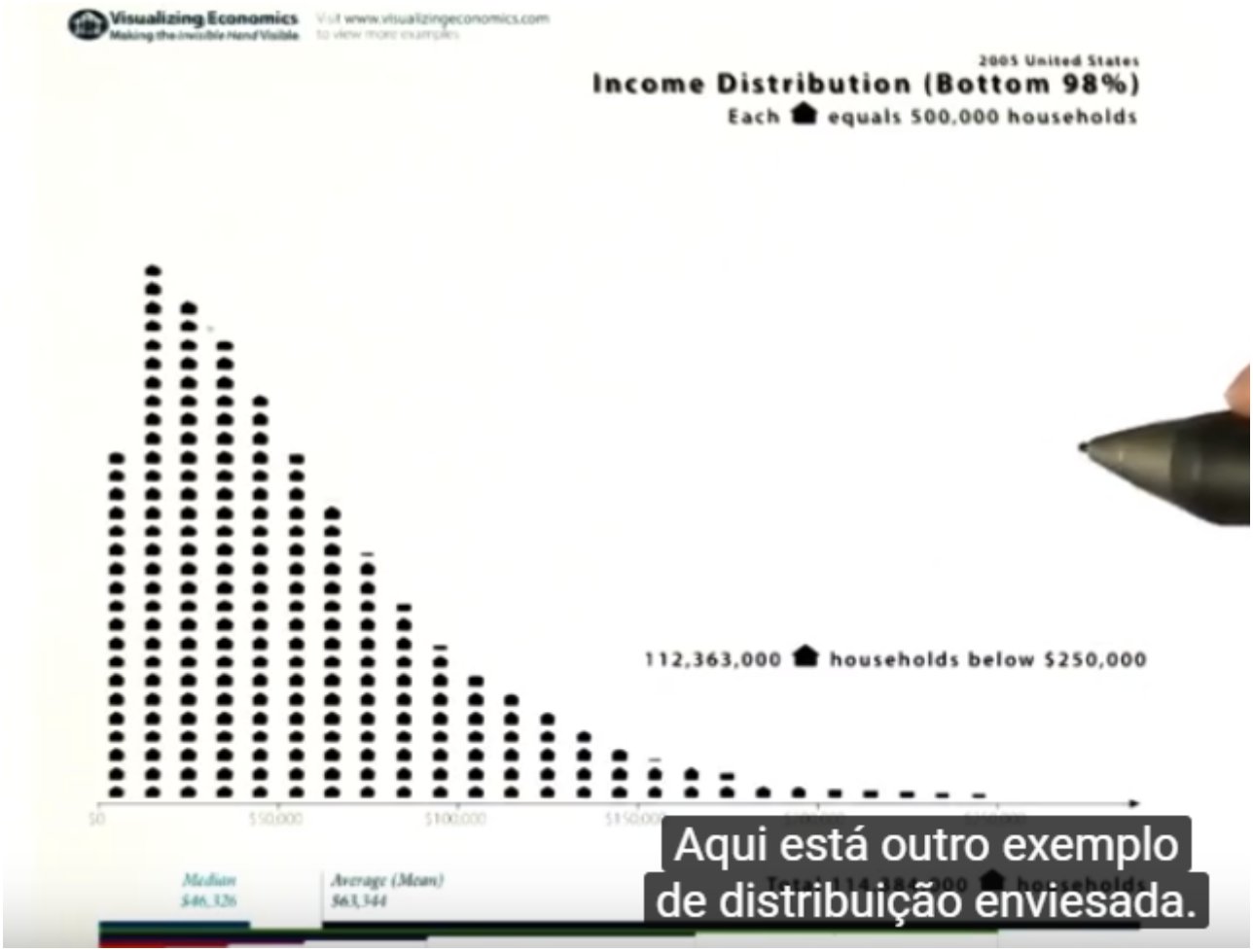
Histograma

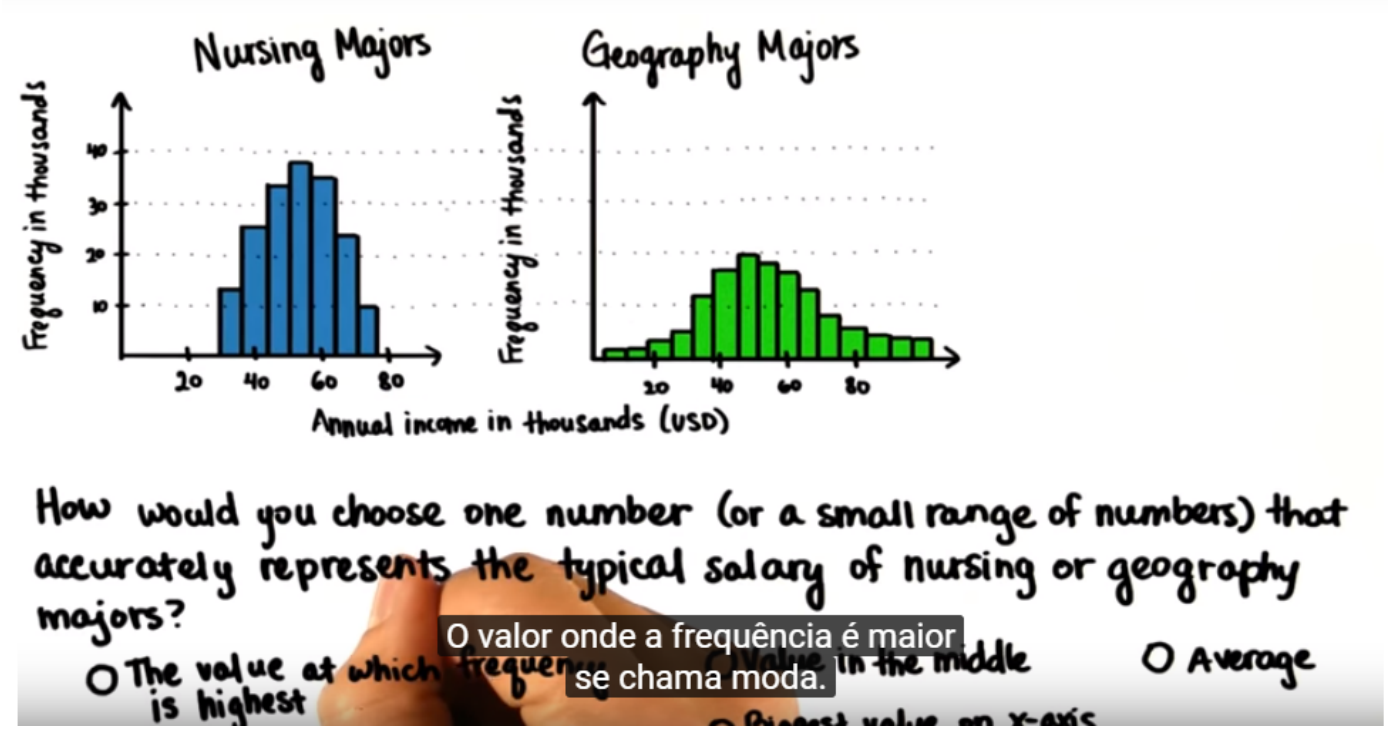
- Frequencia: Variável eixo Y
- Origem: Interseção dos eixos, coordenadas cartesianas (0,0)

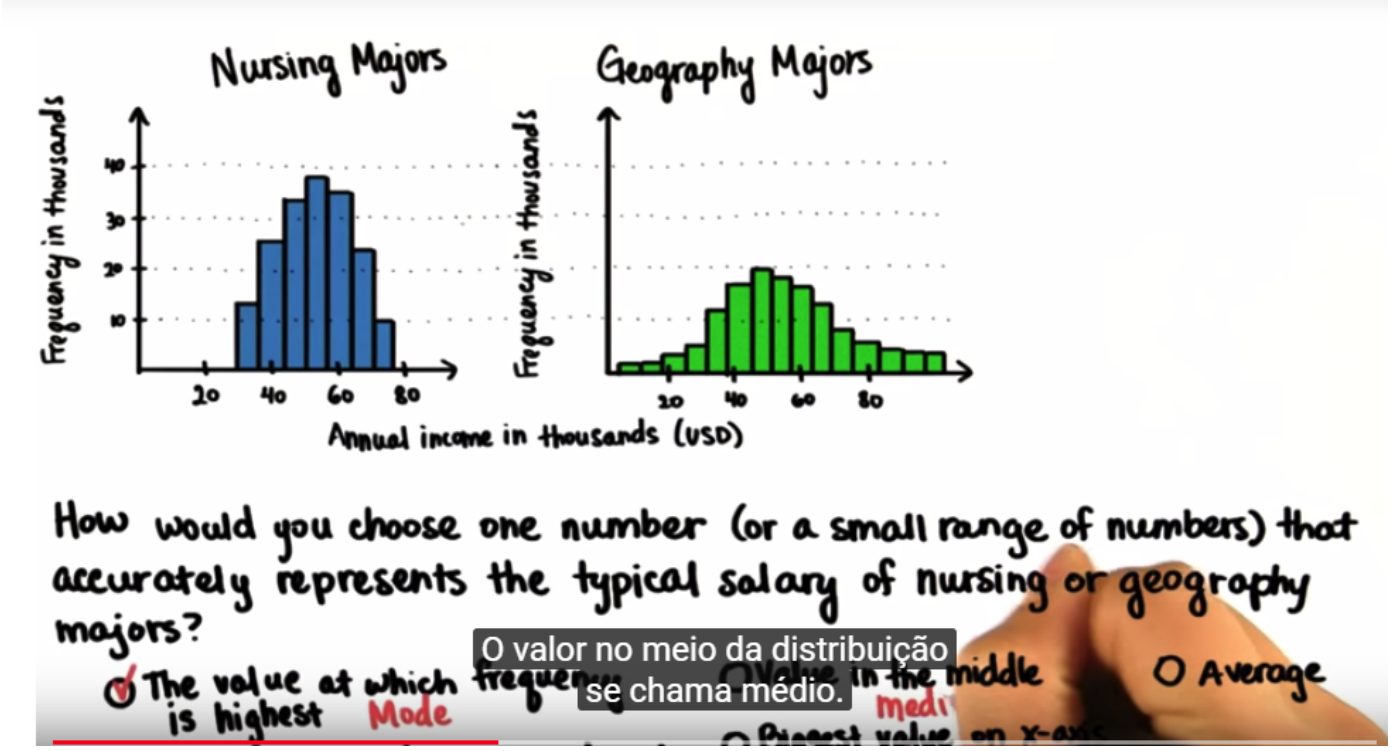
Em um histograma a variável do eixo X é numérica e quantitativa

Em um grafico de barras o eixo X é categórico e qualitativo


HISTOGRAM
x-axis: Numerical/Quantitative
Sample of Udacity Students' ages
Frequency
Age
BAR GRAPH
x-axis: Categorical/qualitative
Sample of Udacity Students' Continents
Frequency (%)
30
25
20
15
10
5
é categórica ou qualitativa.


Visualizing Economics
2005 United States
Income Distribution (Bottom 98%)
Each house equals 500,000 households
112,363,000 households below $250,000
$0
$50,000
$100,000
$150,000
Median
$46,326
Average (Mean)
$63,344
Aqui está outro exemplo
de distribuição enviesada.



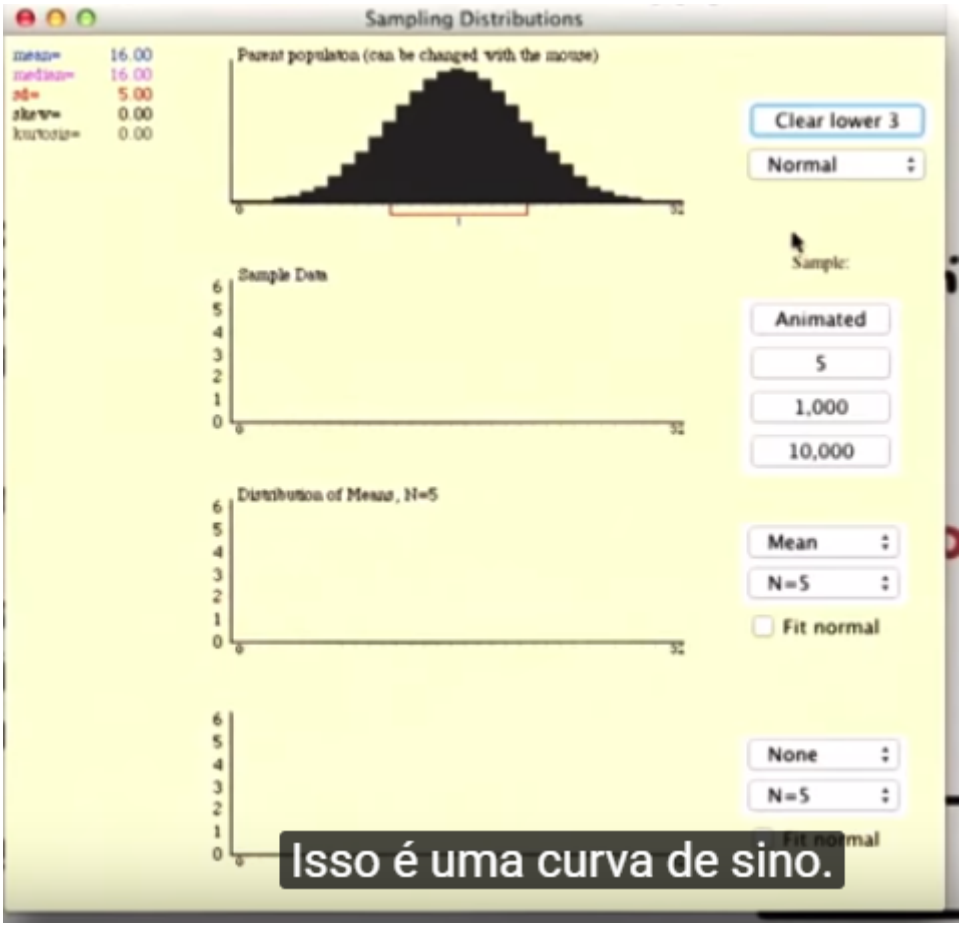


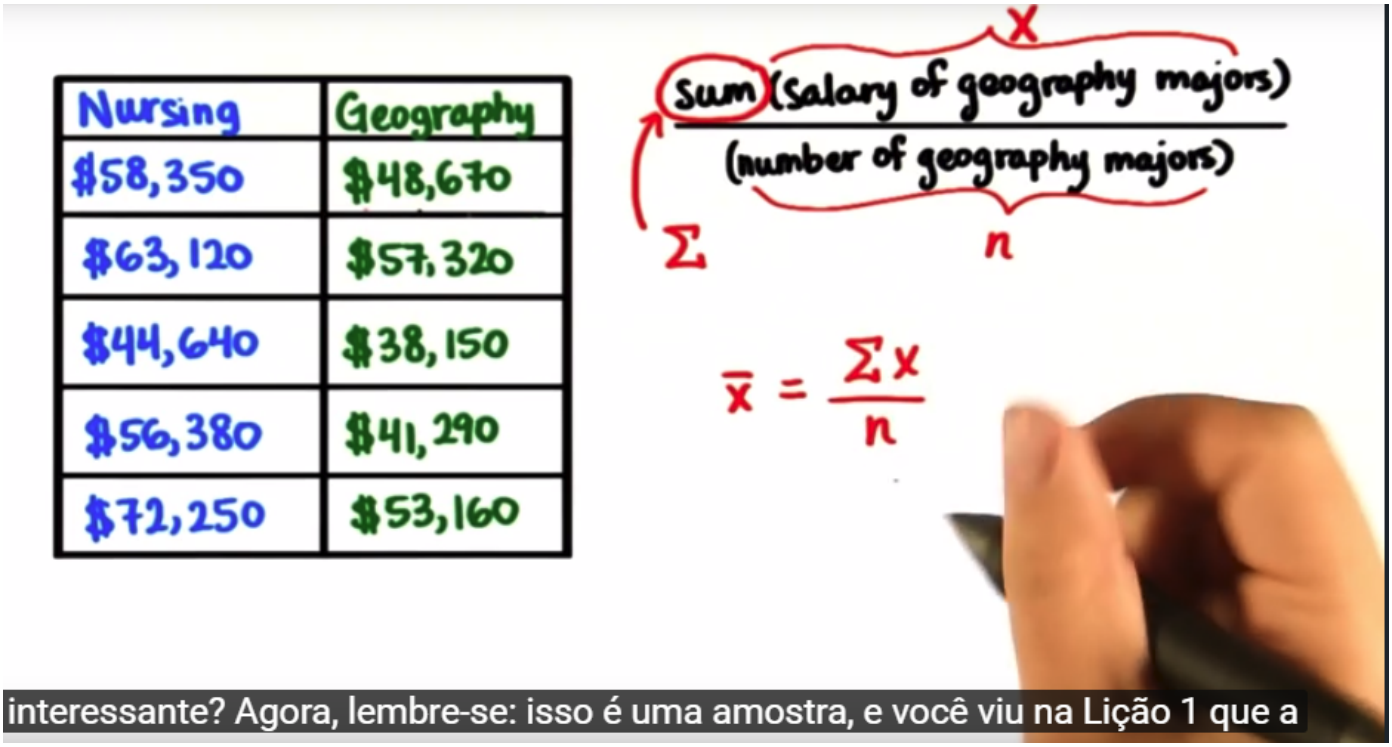
Quando a distribui;'ao possui varias MODAS, neste caso chamados de **DISTRIBUIÇÃO UNIFORME**

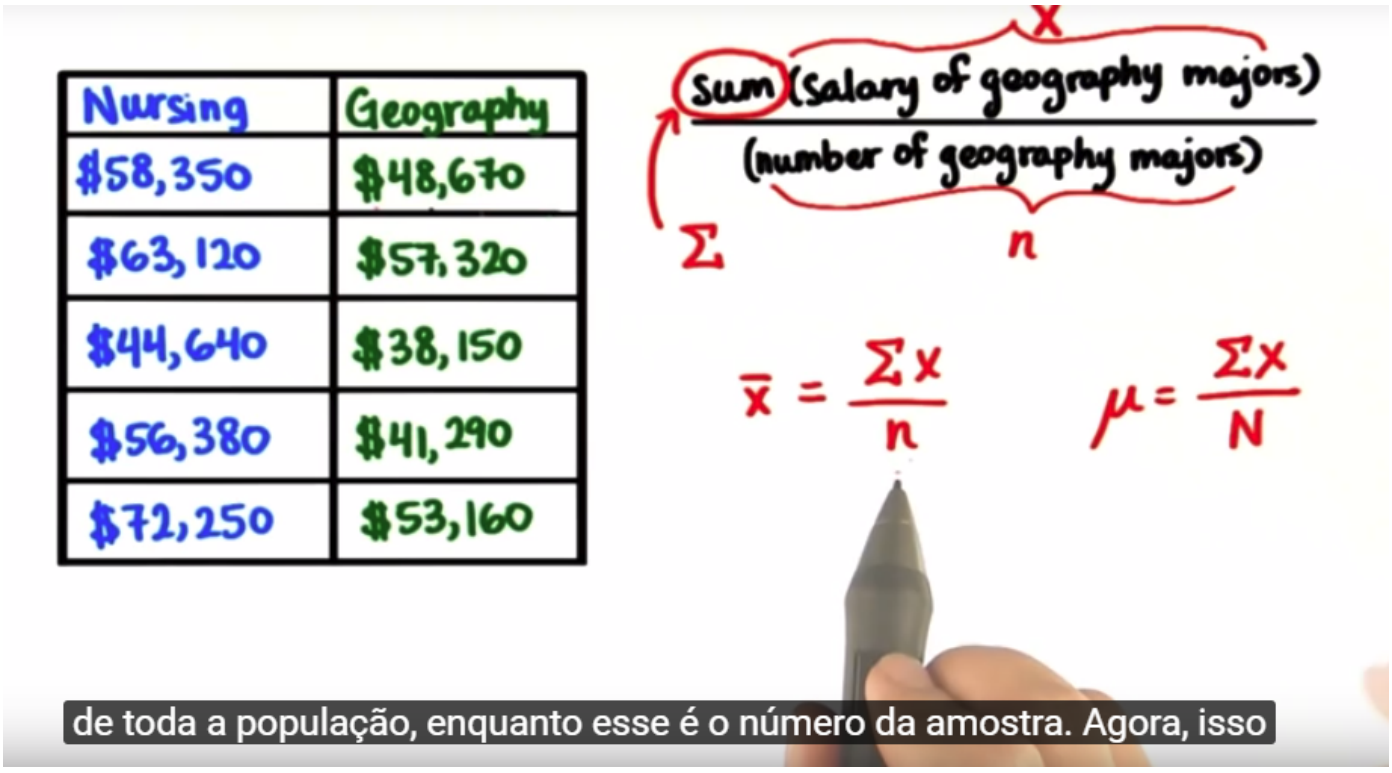
BIMODULAR

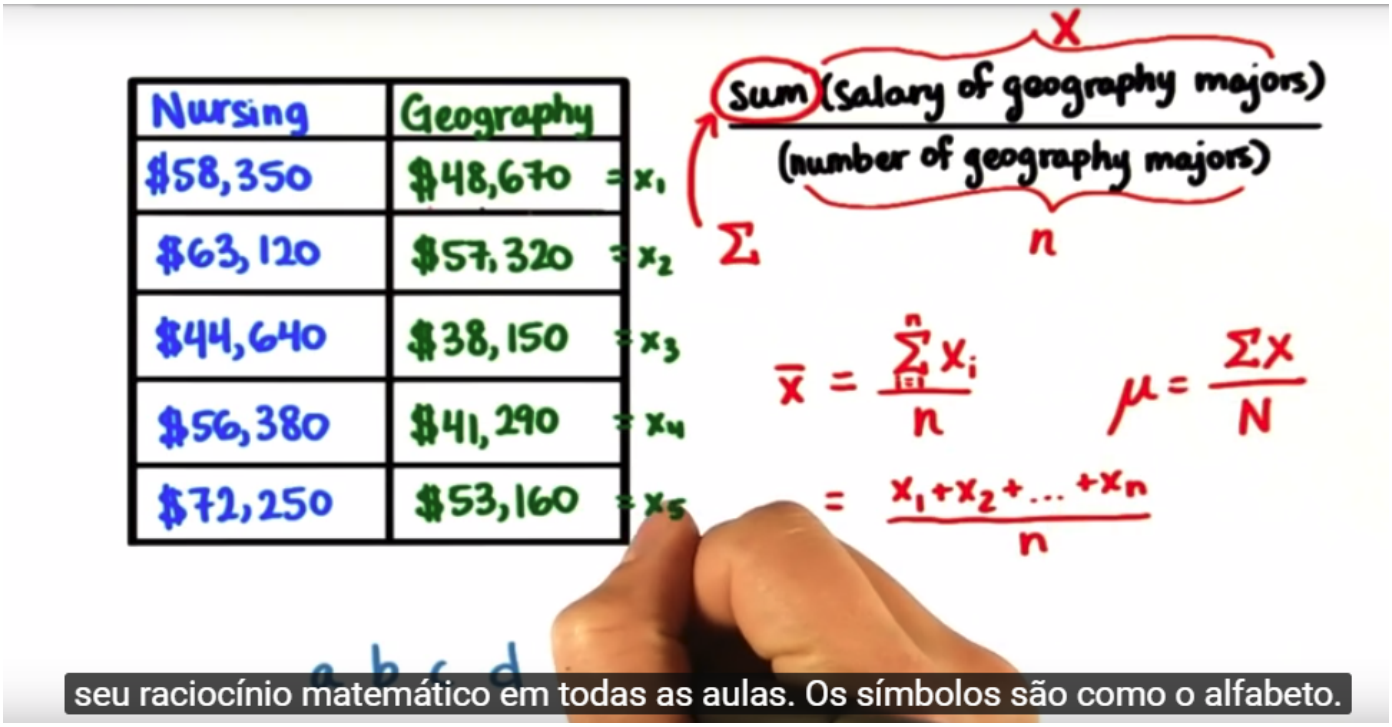
Answer Remember, the mode occurs on the X-axis, so you are looking for whatever value has the highest frequency.

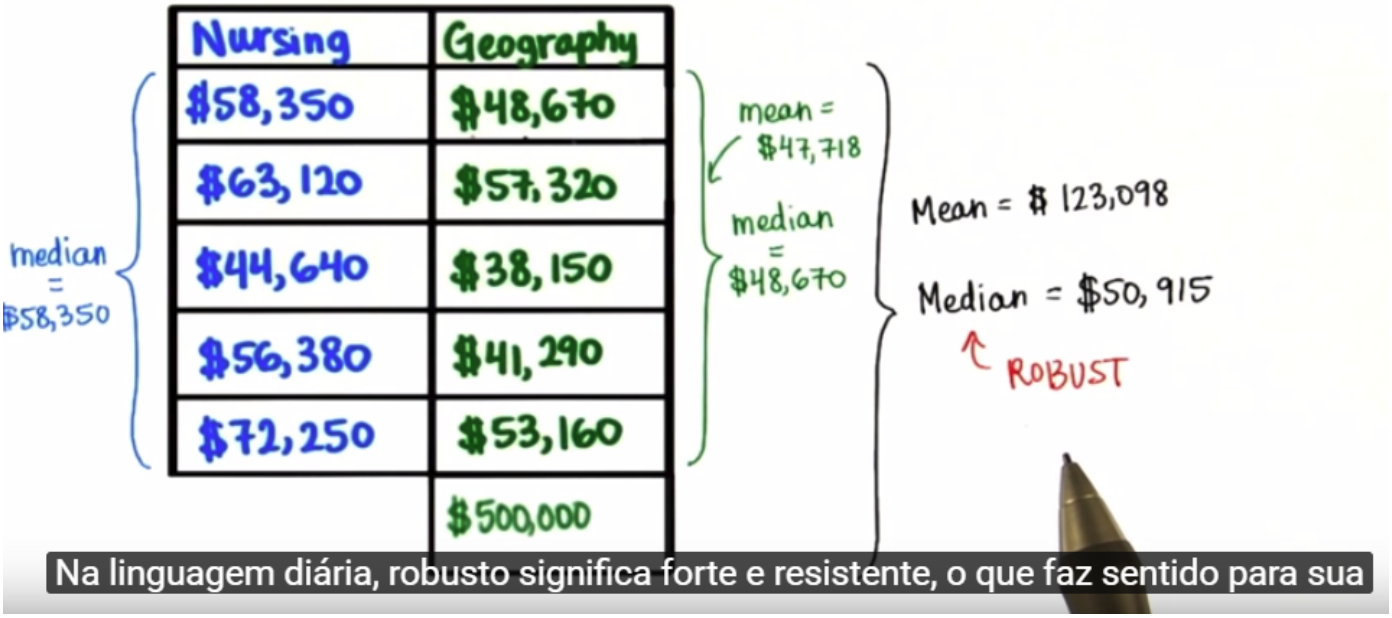
The numbers 7,000 and 1,000 are the actual frequencies. The mode, itself, is "Plain.

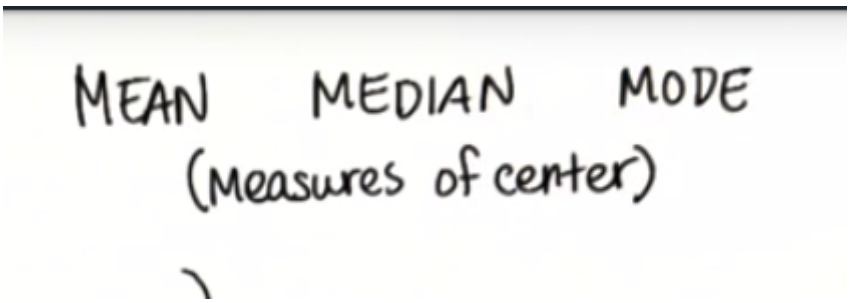



Nursing
Geography
$58,350
$48,670
$63,120
$57,320
$44,640
$38,150
$56,380
$41,290
$72,250
$53,160
Sum (Salary of geography majors)
(number of geography majors)
x
n
Σ
x̄ = Σx / n
interessante? Agora, lembre-se: isso é uma amostra, e você viu na Lição 1 que a


Nursing
Geography
$58,350
$48,670
$63,120
$57,320
$44,640
$38,150
$56,380
$41,290
$72,250
$53,160
Sum (Salary of geography majors)
(number of geography majors)
x
n
Σ
x̄ = Σx / n
μ = Σx / N
de toda a população, enquanto esse é o número da amostra. Agora, isso



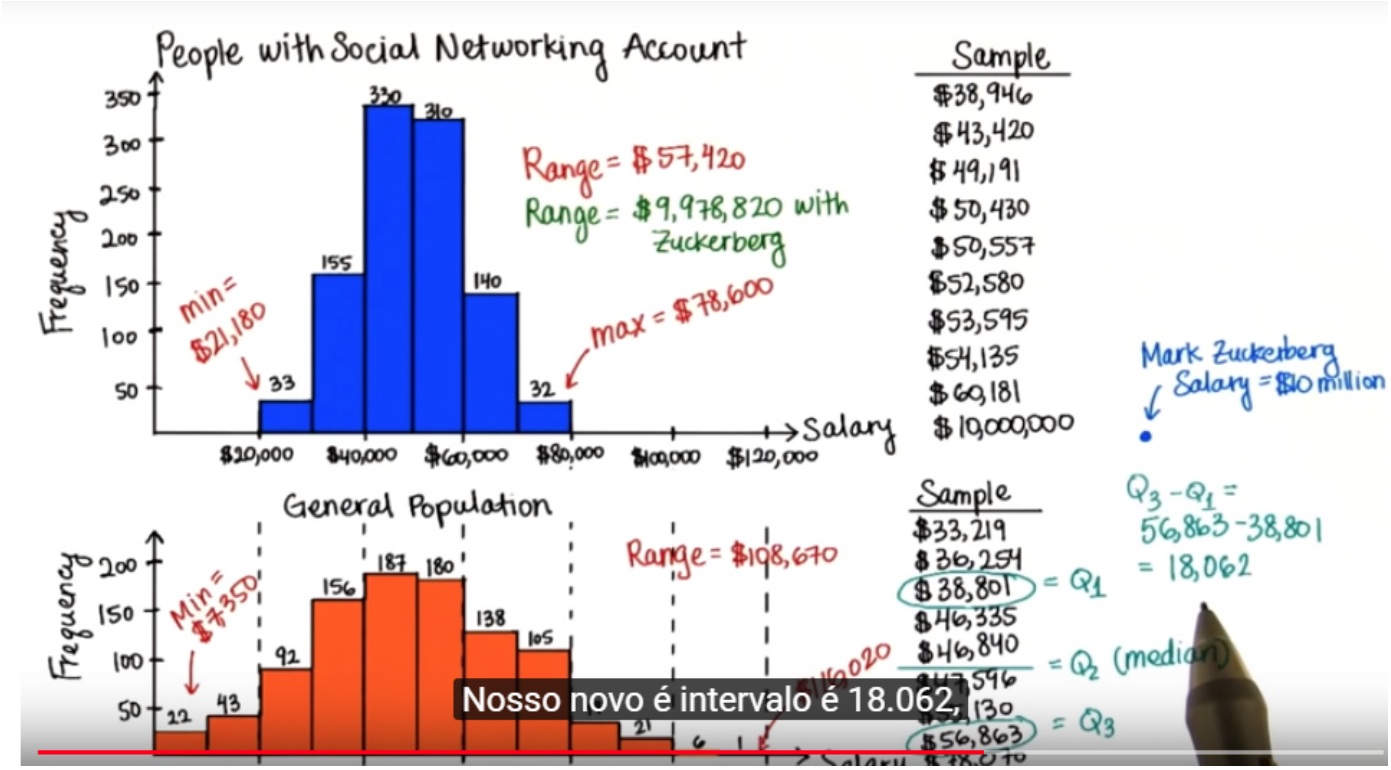




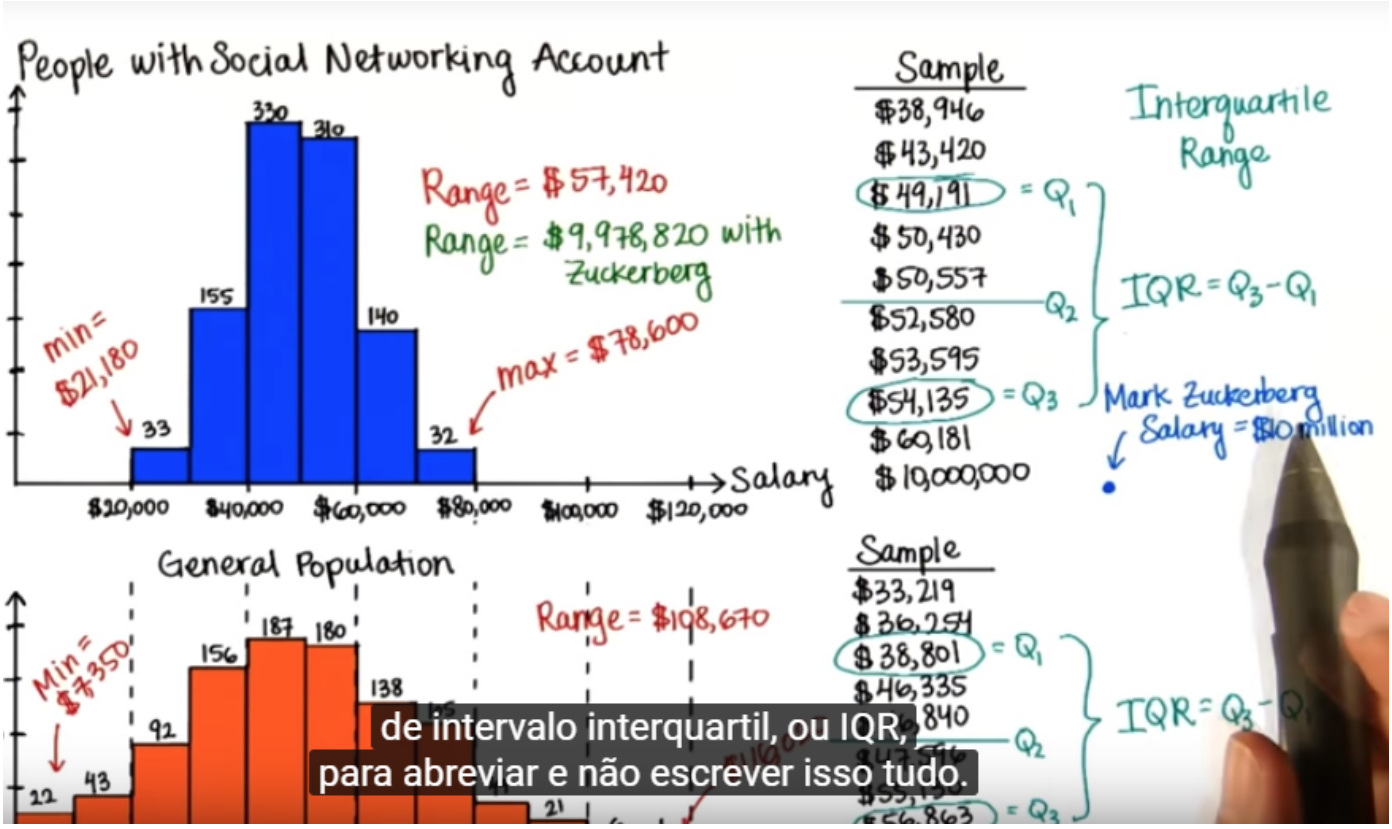
Como visto em aula, a média pode não descrever o centro devido a um outlier, as vezes a moda também não descreveo centro e a mediana não abrange todos os pontos de dados.

Muitas vezes a mediana pode ser mais adequada para trabalhar com graficos altamente enviesados

Nesta etapa é calculado os quartis, para isso se divide os valores da amostra em 2 partes e depois cada valor mediano de cada conjunto são os Q1 = Primeiro quartis e Q3 = Terceiro quartis



Para calcular a nova distribuição faça Q3 - Q1, assim vc terá o que chamamos de intervalo interquartil

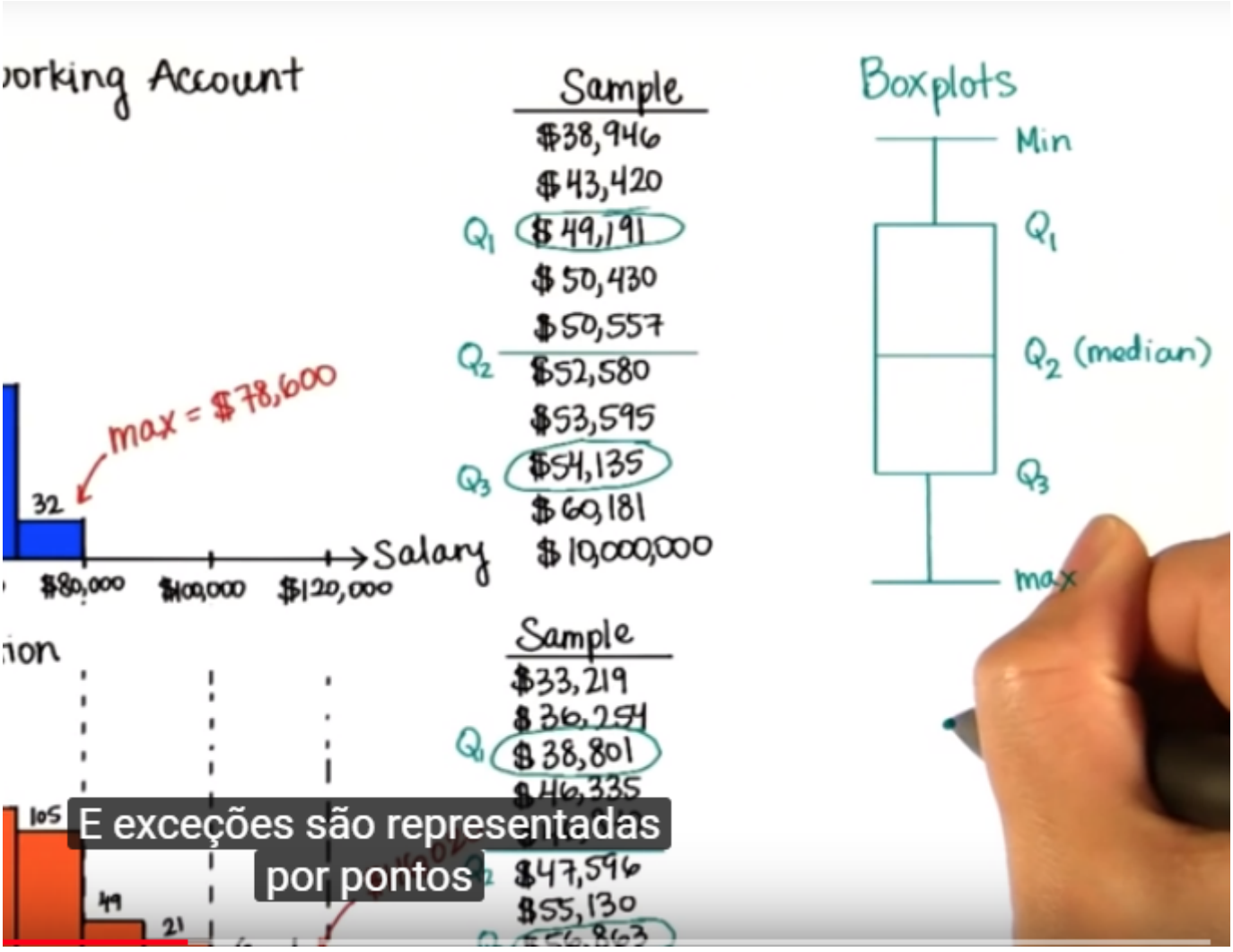


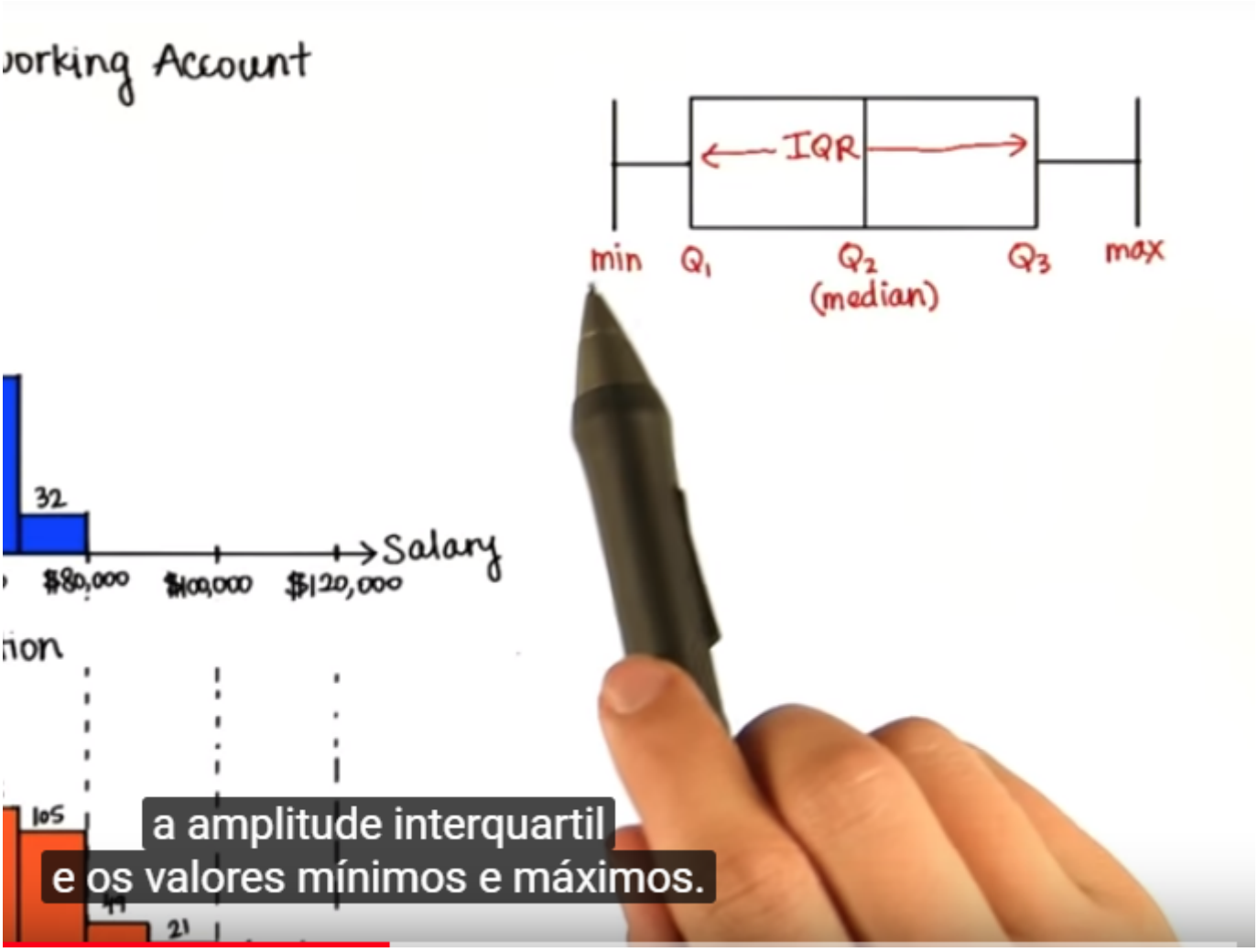
Outlier $< Q_1 - 1.5(IQR)$

$> Q_3 + 1.5(IQR)$

What values do you think are outliers for this dataset?

- [ ] $60,000
- [x] $80,000
- [x] $100,000
- [x] $200,000

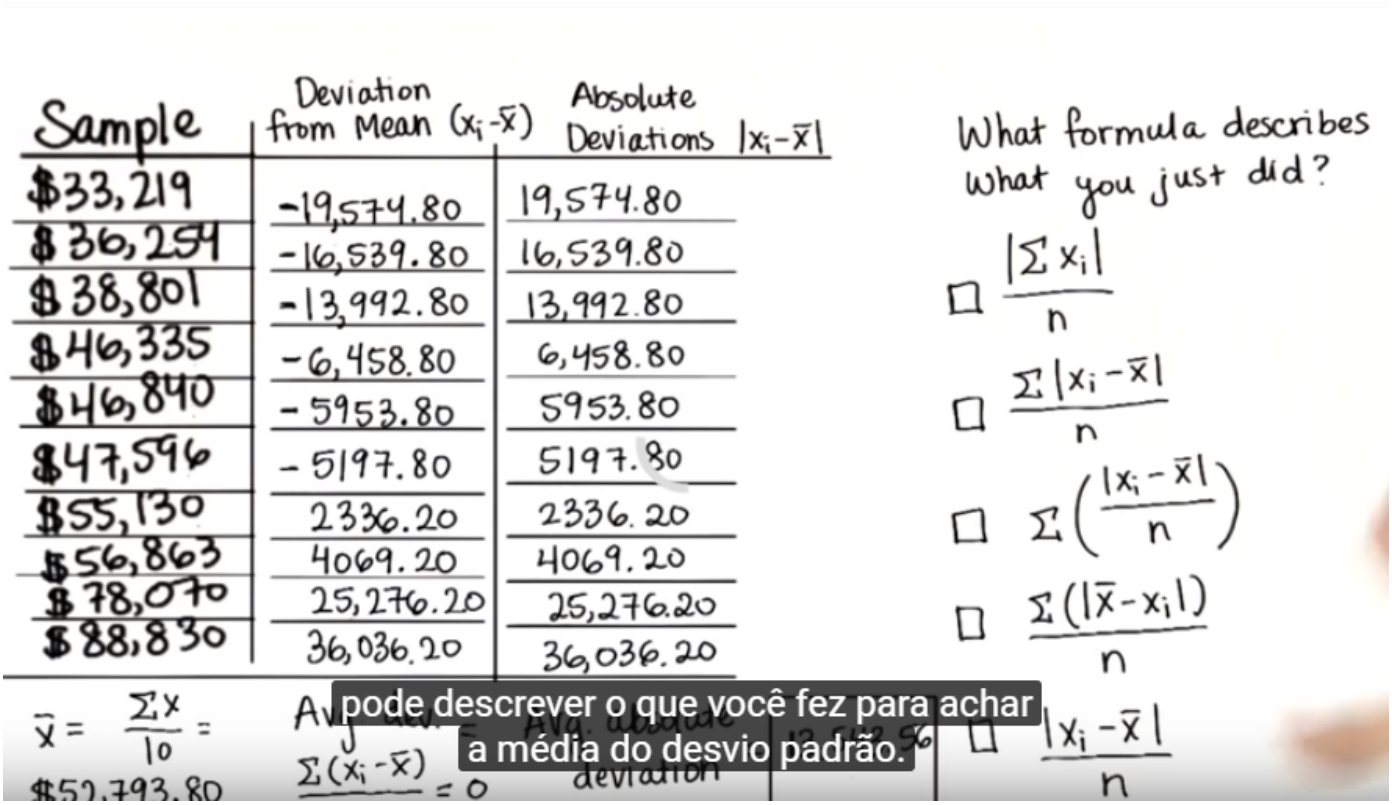




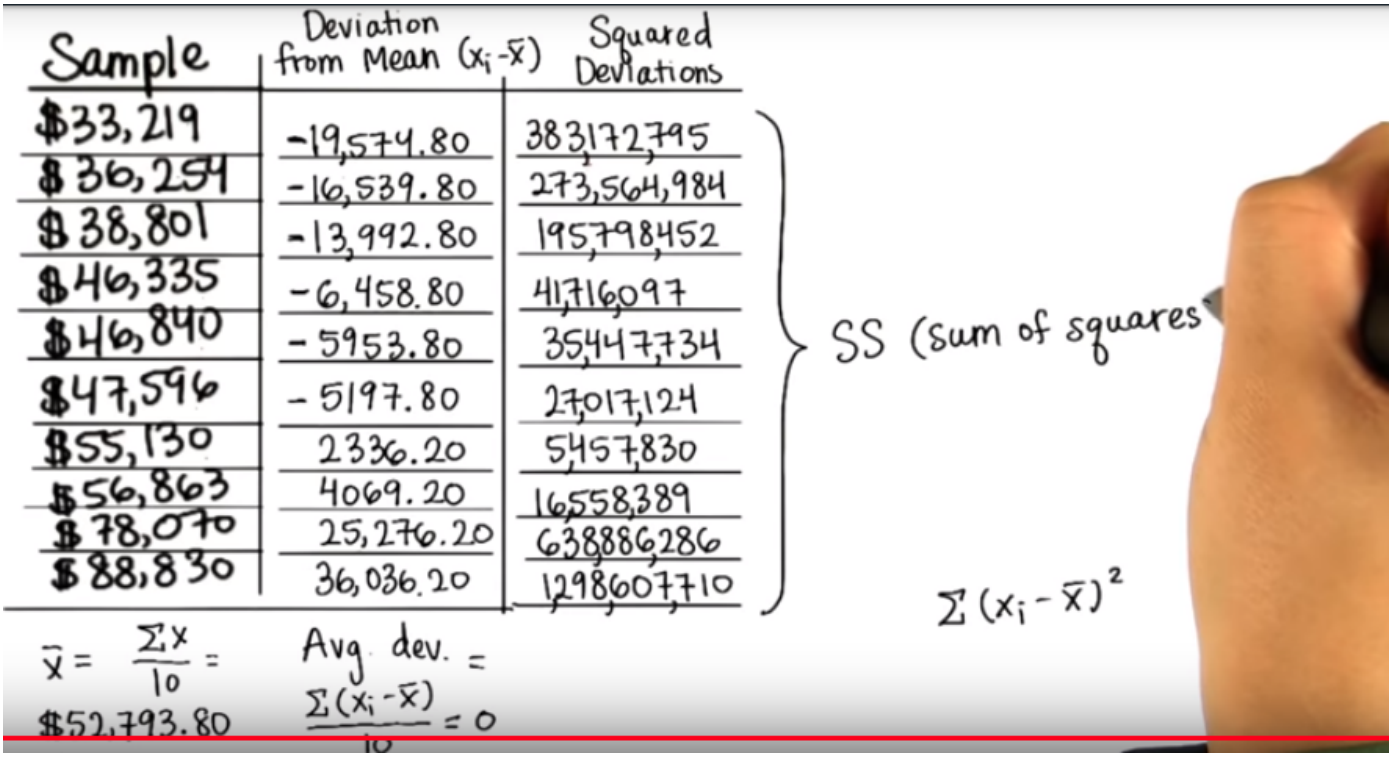
Sample	Deviation from Mean $(x_i - \bar{x})$
$33,219	-19,574.80
$36,259	-16,539.80
$38,801	-13,992.80
$46,335	-6,458.80
$46,840	-5953.80
$47,596	-5197.80
$55,130	2336.20
$56,863	4069.20
$78,070	25,276.20
$88,830	36,036.20

$\bar{x} = \frac{\sum x}{10} = \$52,793.80$

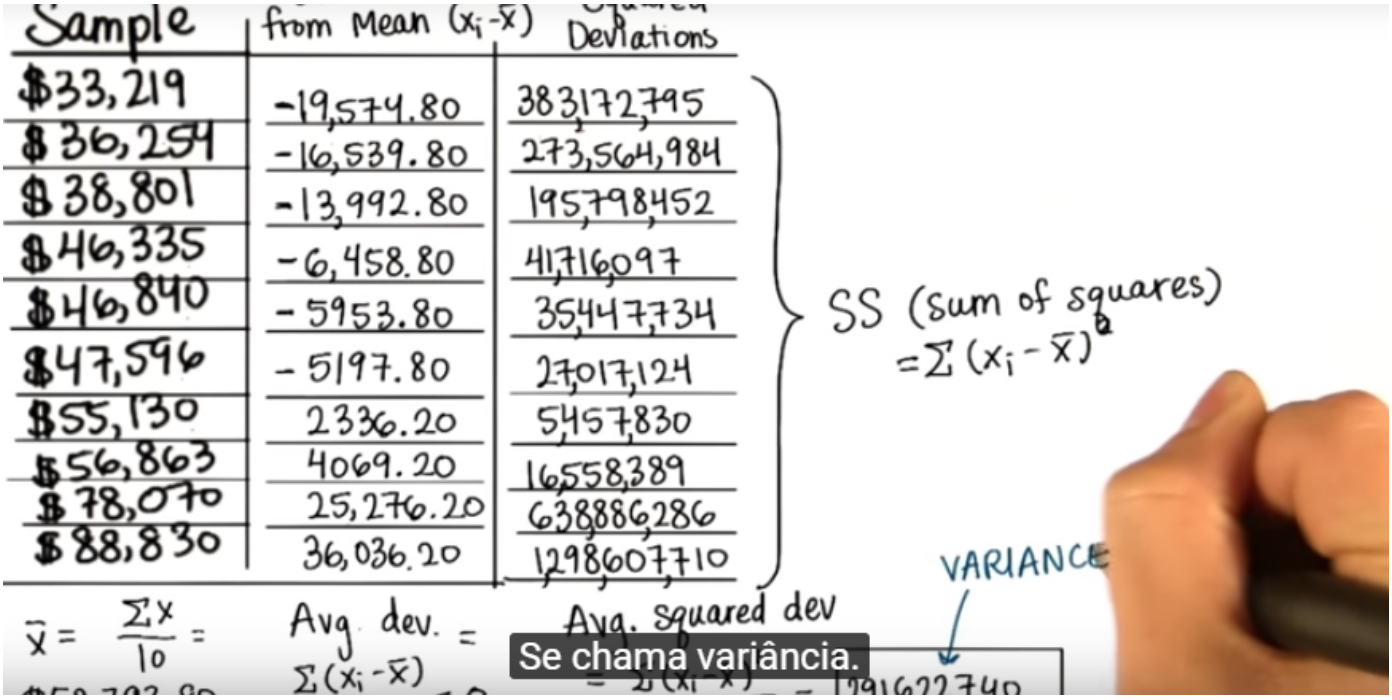
ao invés de fazer o oposto,
x barra menos xi.

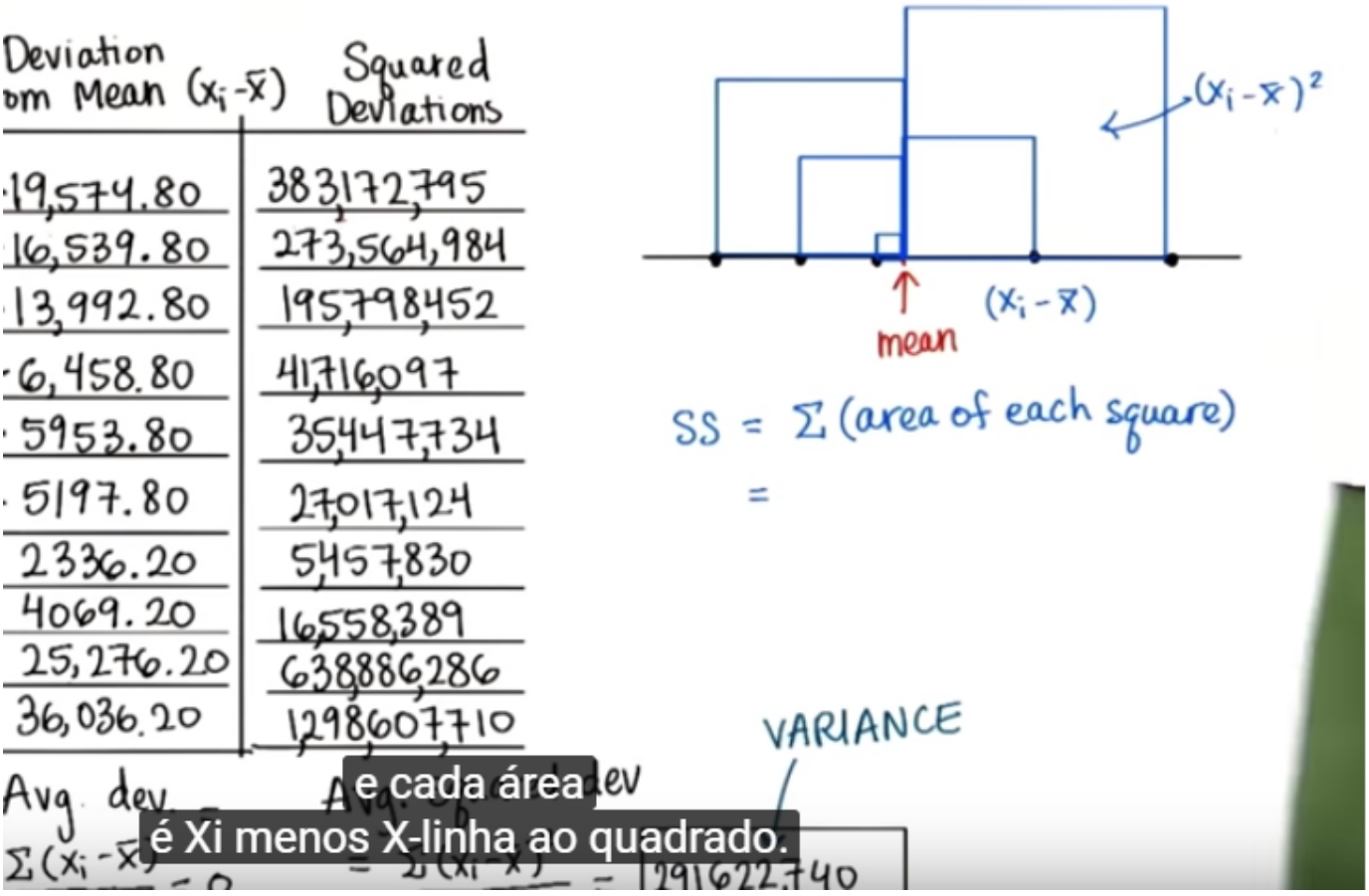


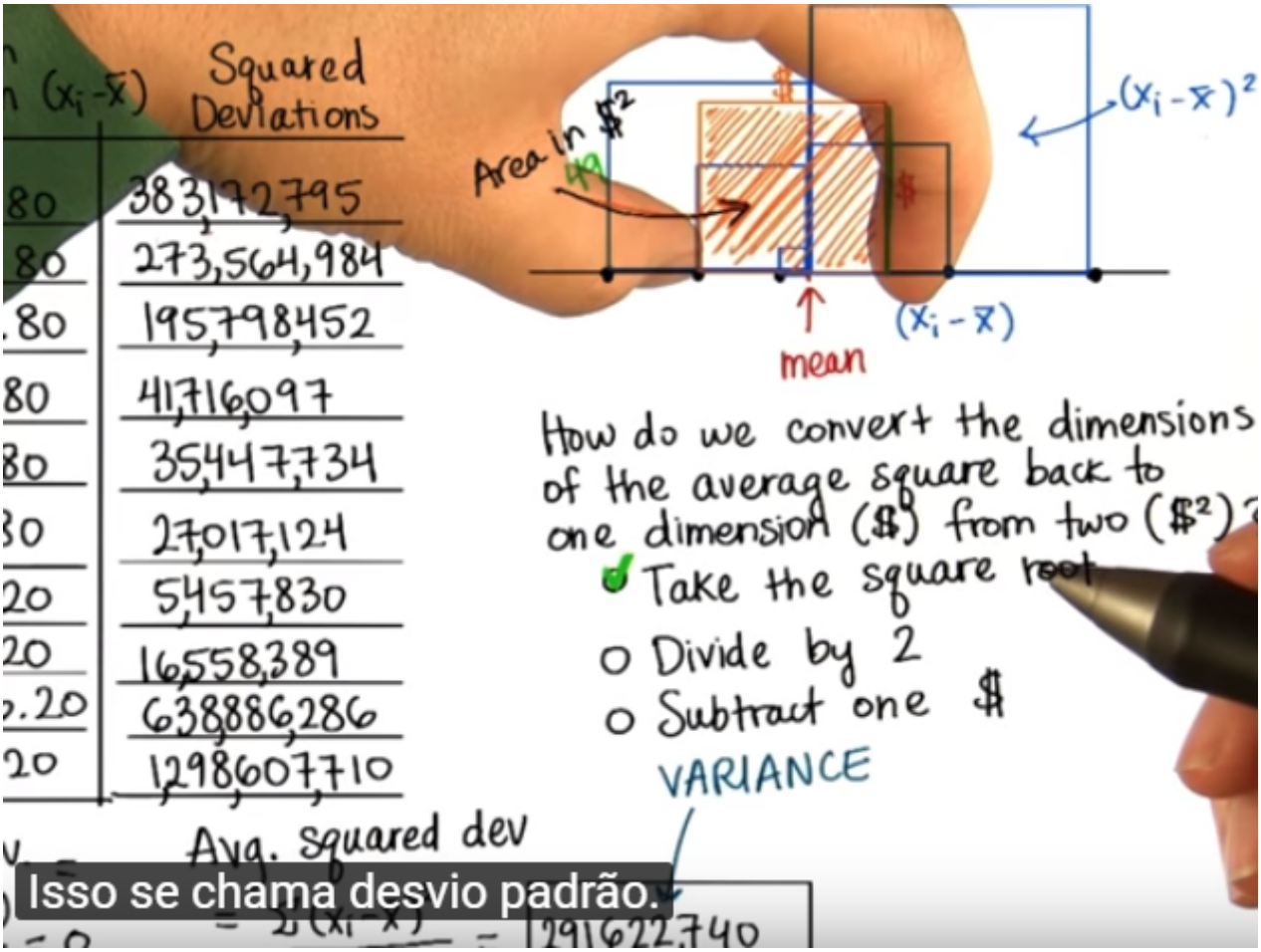
pode descrever o que você fez para achar a média do desvio padrão.



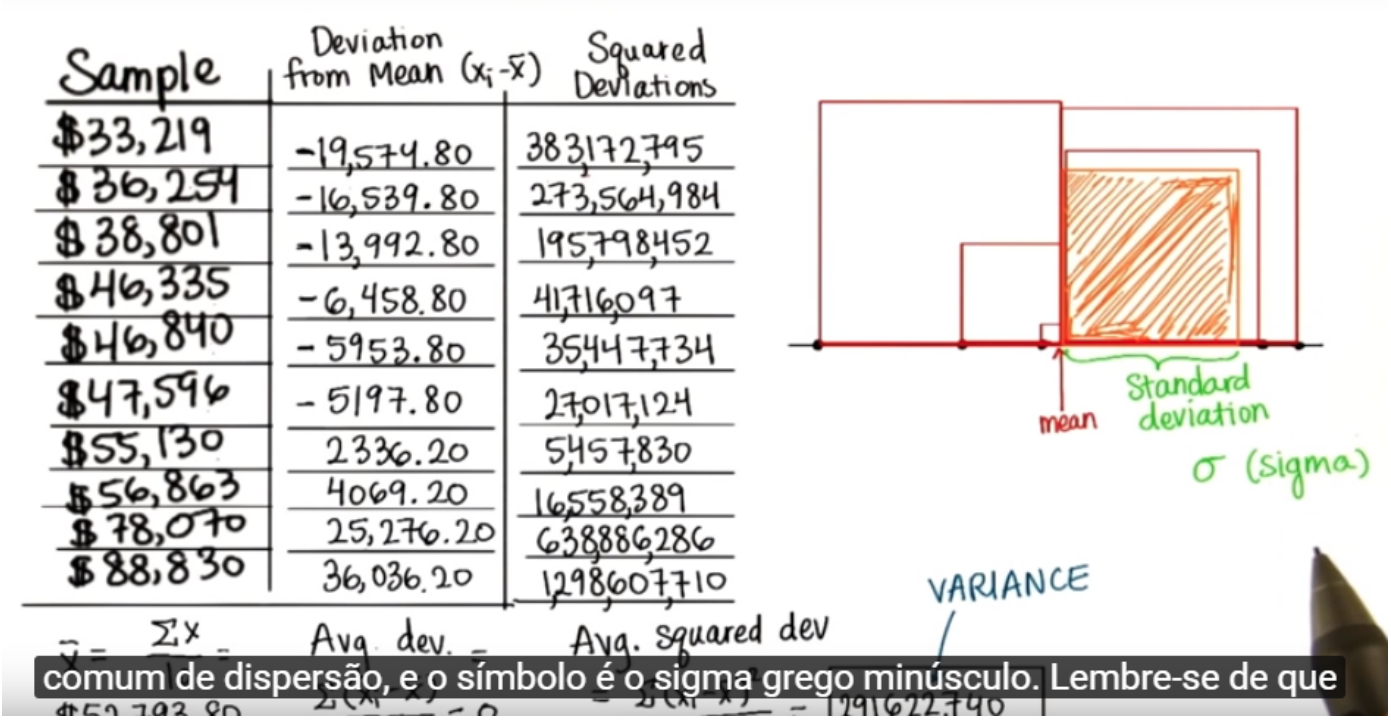
A variancia é a soma dos desvios ao quadrado dividivo pelo total de itens (média)

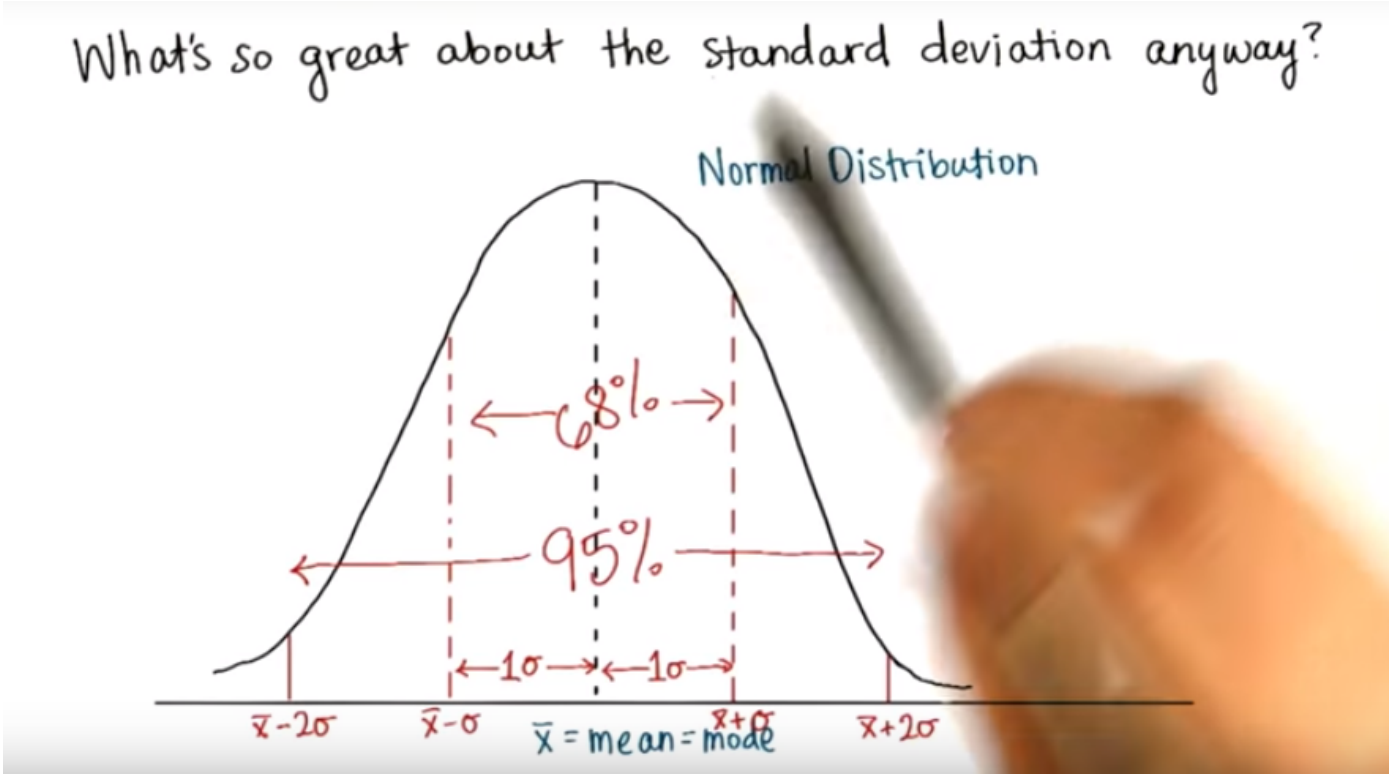

Sample
from Mean ($x_i - \bar{x}$)
Deviations
$33,219 | -19,574.80 | 383,172,795
$36,259 | -16,539.80 | 273,564,984
$38,801 | -13,992.80 | 195,798,452
$46,335 | -6,458.80 | 41,716,097
$46,840 | -5953.80 | 35,447,734
$47,596 | -5197.80 | 27,017,124
$55,130 | 2336.20 | 5,457,830
$56,863 | 4069.20 | 16,558,389
$78,070 | 25,276.20 | 638,886,286
$88,830 | 36,036.20 | 1,298,607,710
SS (sum of squares) $= \sum (x_i - \bar{x})^2$
$\bar{x} = \frac{\sum x}{10} =$
Avg. dev. $= \frac{\sum (x_i - \bar{x})}{}$
Avg. squared dev
VARIANCE
Se chama variância.

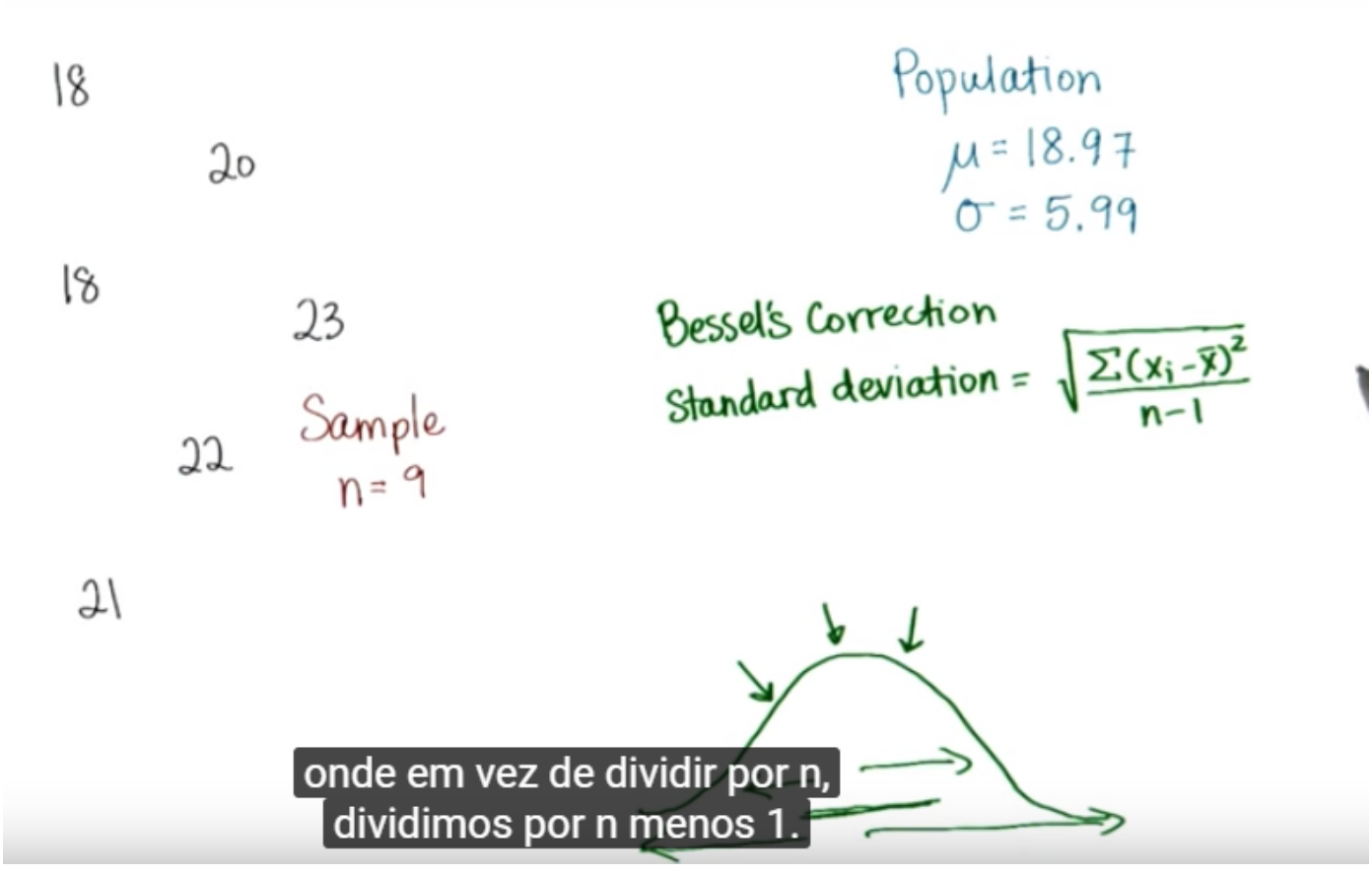

Deviation
om Mean ($x_i - \bar{x}$)
Squared
Deviations
-19,574.80 | 383,172,795
-16,539.80 | 273,564,984
-13,992.80 | 195,798,452
-6,458.80 | 41,716,097
-5953.80 | 35,447,734
-5197.80 | 27,017,124
2336.20 | 5,457,830
4069.20 | 16,558,389
25,276.20 | 638,886,286
36,036.20 | 1,298,607,710
$(x_i - \bar{x})^2$
$(x_i - \bar{x})$
mean
SS = $\sum$ (area of each square)
=
Avg. dev.
$\sum (x_i - \bar{x})$
VARIANCE
e cada área
é Xi menos X-linha ao quadrado.

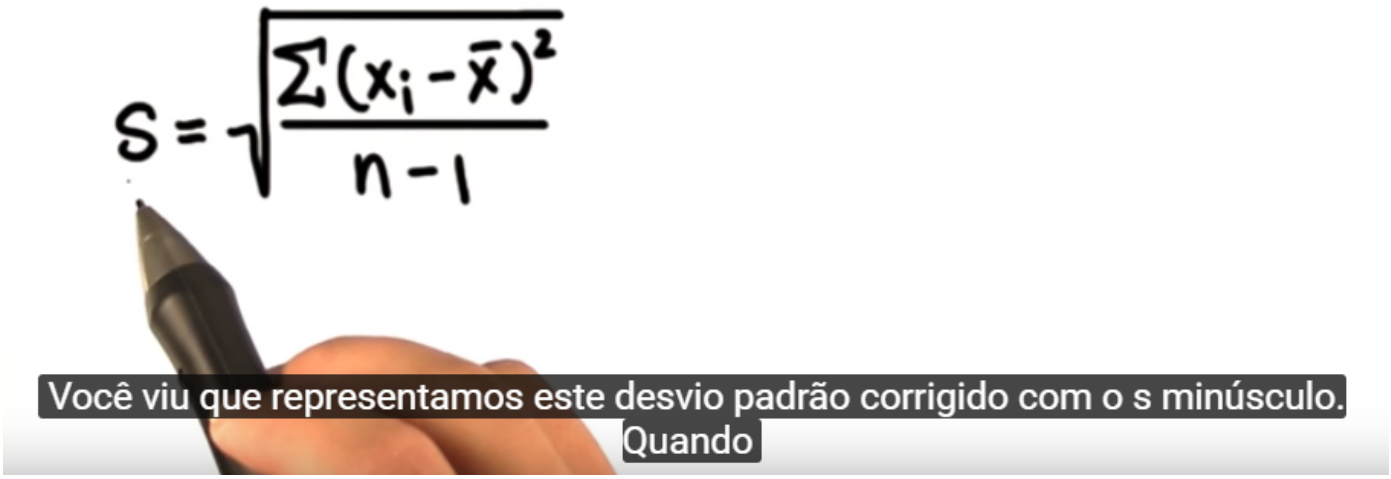


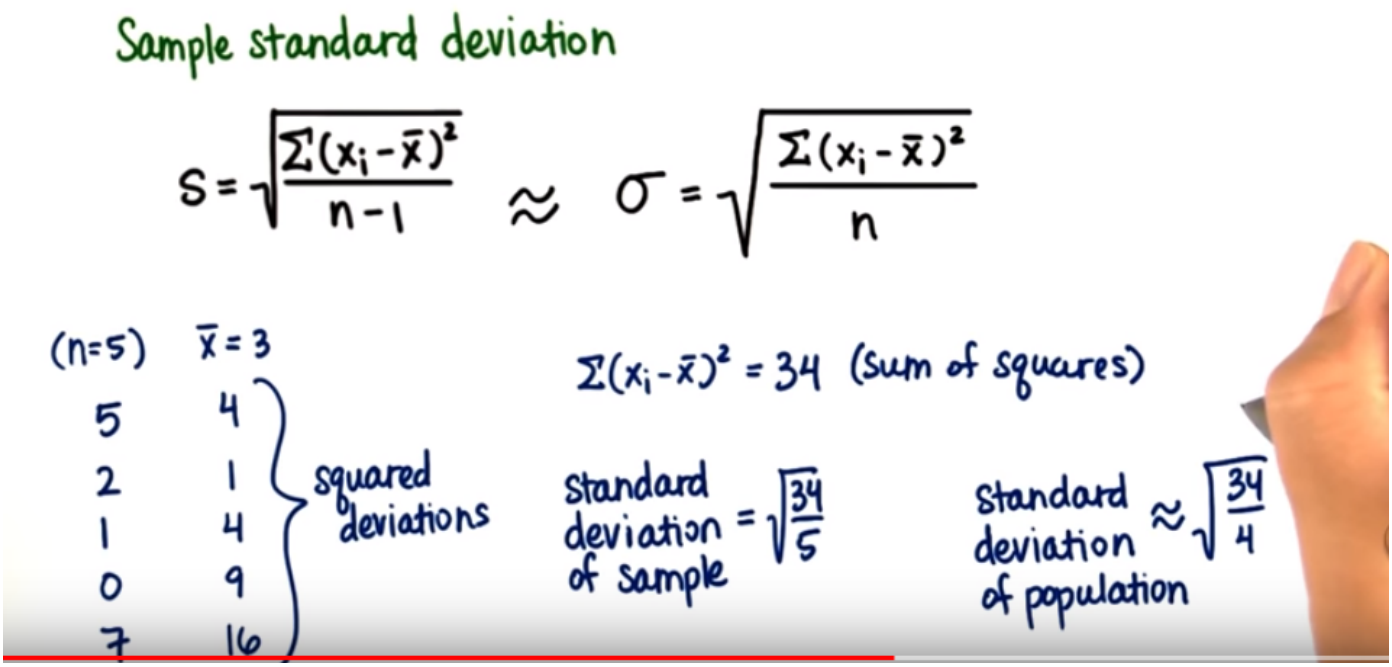
Desvio Padrão = Variancia ao quadrado



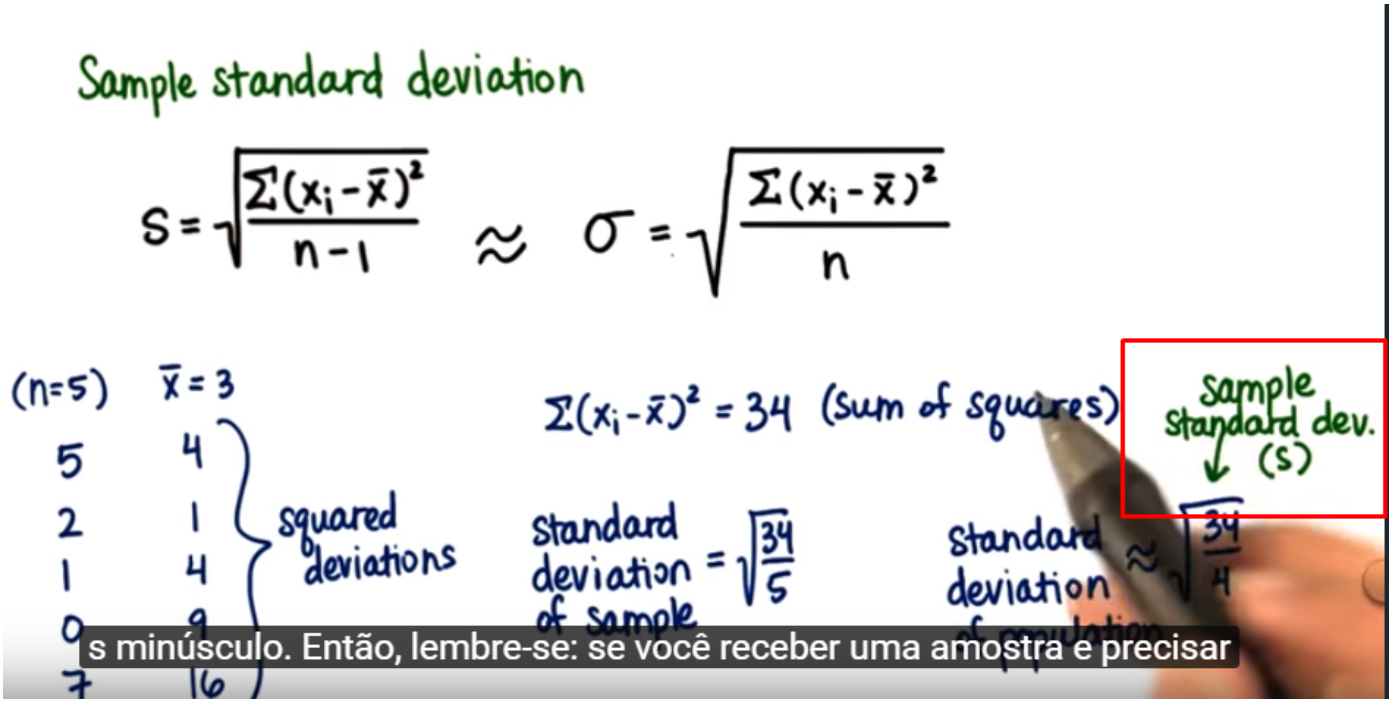

What's so great about the standard deviation anyway?
Normal Distribution
68%
95%
1σ
1σ
x̄-2σ
x̄-σ
x̄+σ
x̄+2σ
x̄ = mean = mode


18
20
18
23
22
Sample
n = 9
21
Population
μ = 18.97
σ = 5.99
Bessel's Correction
Standard deviation = √(Σ(xi - x̄)² / n-1)
onde em vez de dividir por n,
dividimos por n menos 1.



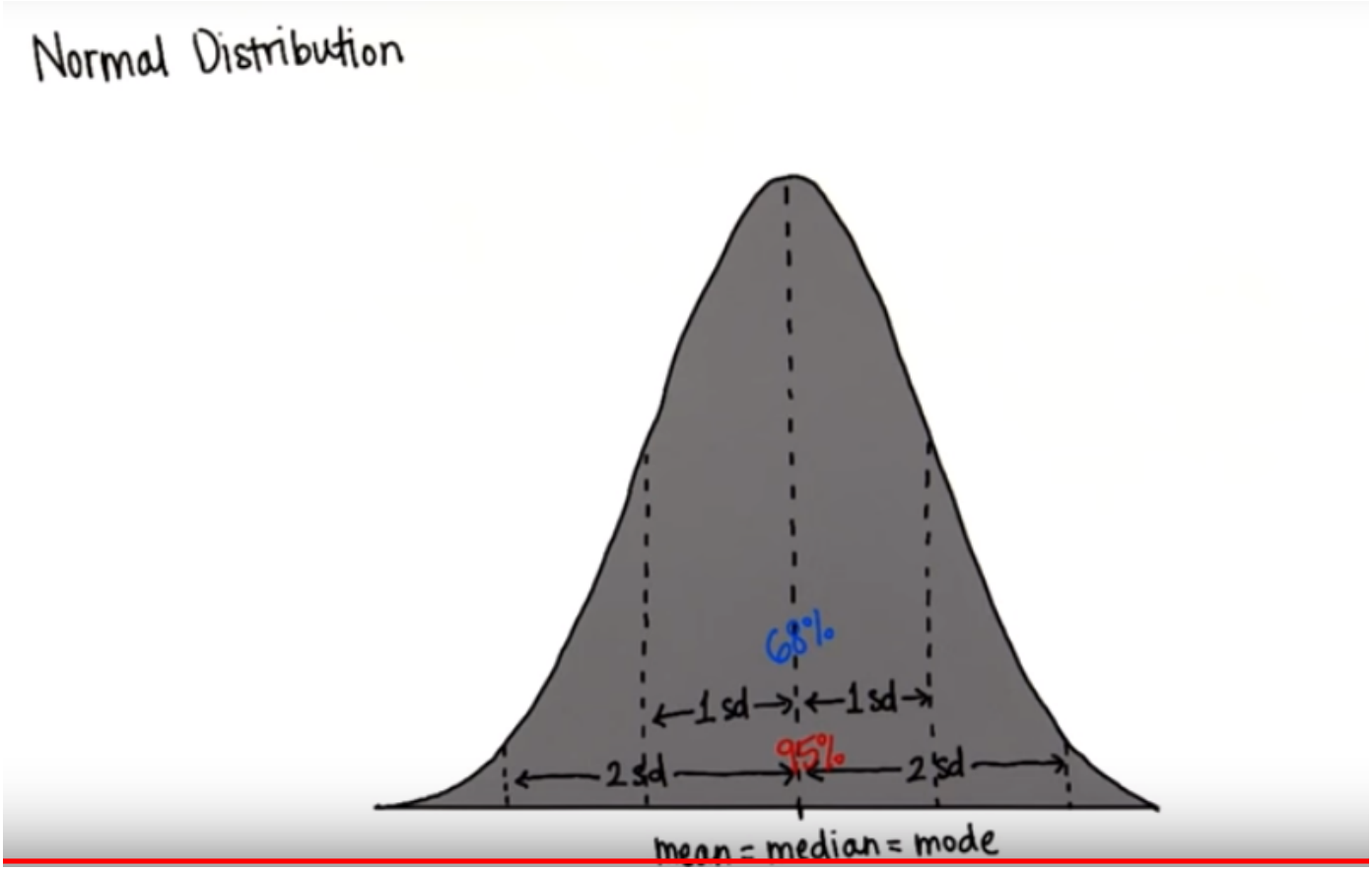


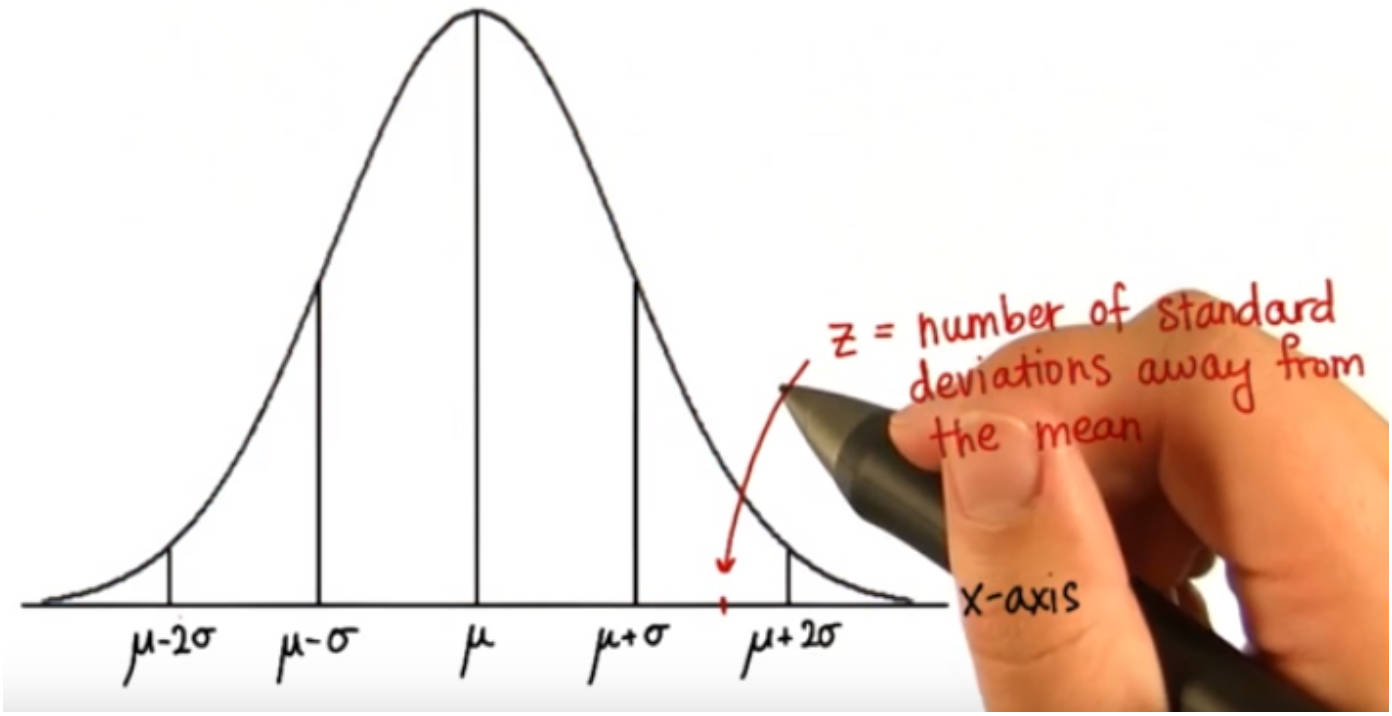
Quando o desvio padrão precisa estar aproximado ao valor da população faça o n-1 para aumentar a distribuição, mas quando não houver esta necessidade faça a formula padrão

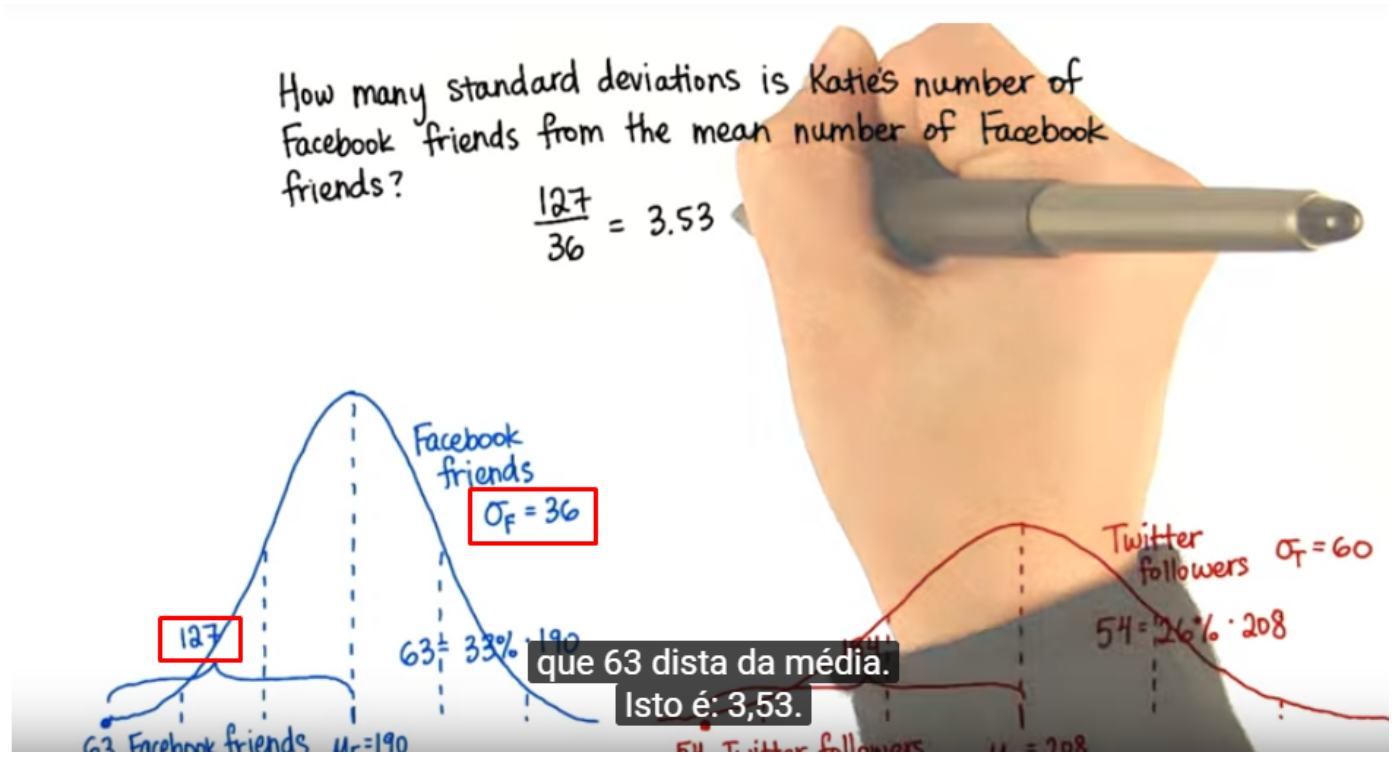


Modelo para calcular um valor % dentro da amostra

Área sempre será 1




z = number of standard deviations away from the mean
x-axis
μ-2σ
μ-σ
μ
μ+σ
μ+2σ


How many standard deviations is Katie's number of Facebook friends from the mean number of Facebook friends?
127/36 = 3.53
Facebook friends
σF = 36
127
63 = 33%
Twitter followers
σT = 60
54 = 26% · 208
que 63 dista da média.
Isto é: 3,53.

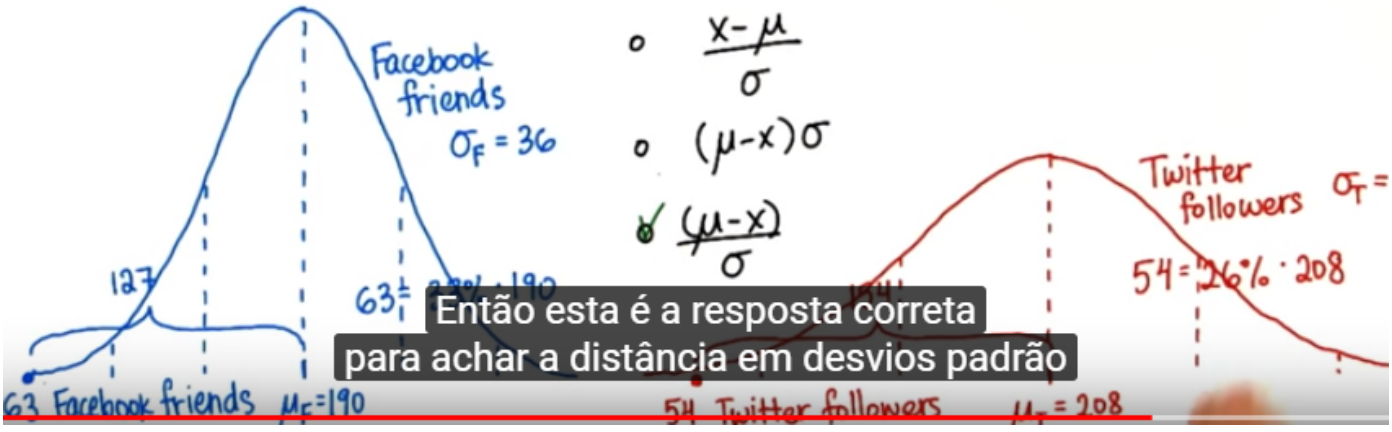
If Andy only uses Twitter and I only use Facebook, can we necessarily say that Andy is more unpopular than me?

- ○ Yes — Why?
- ✓ No

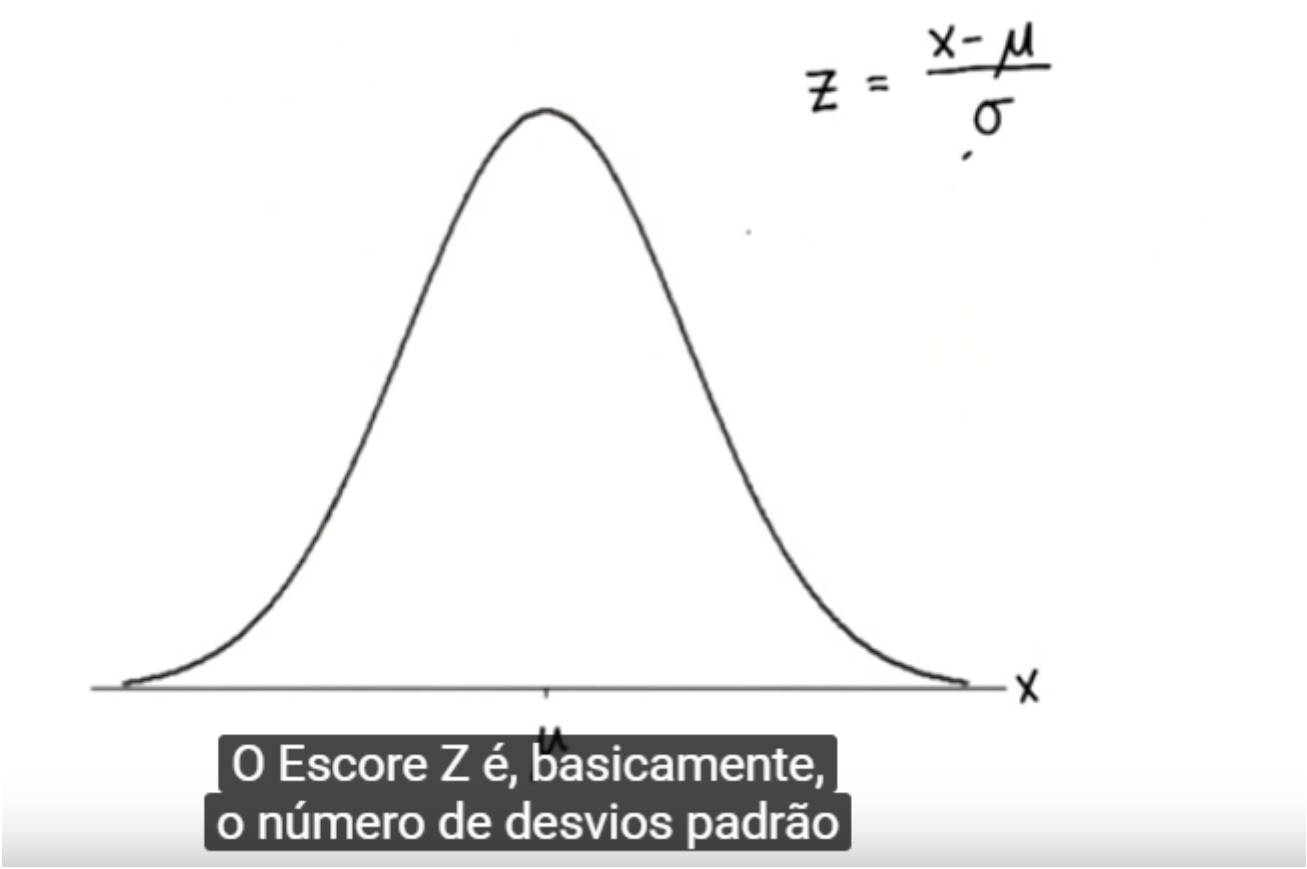
Isso se chama padronizar as distribuições.

What formula describes what you did to find the number of standard deviations each value (let's call it x) is from the mean?

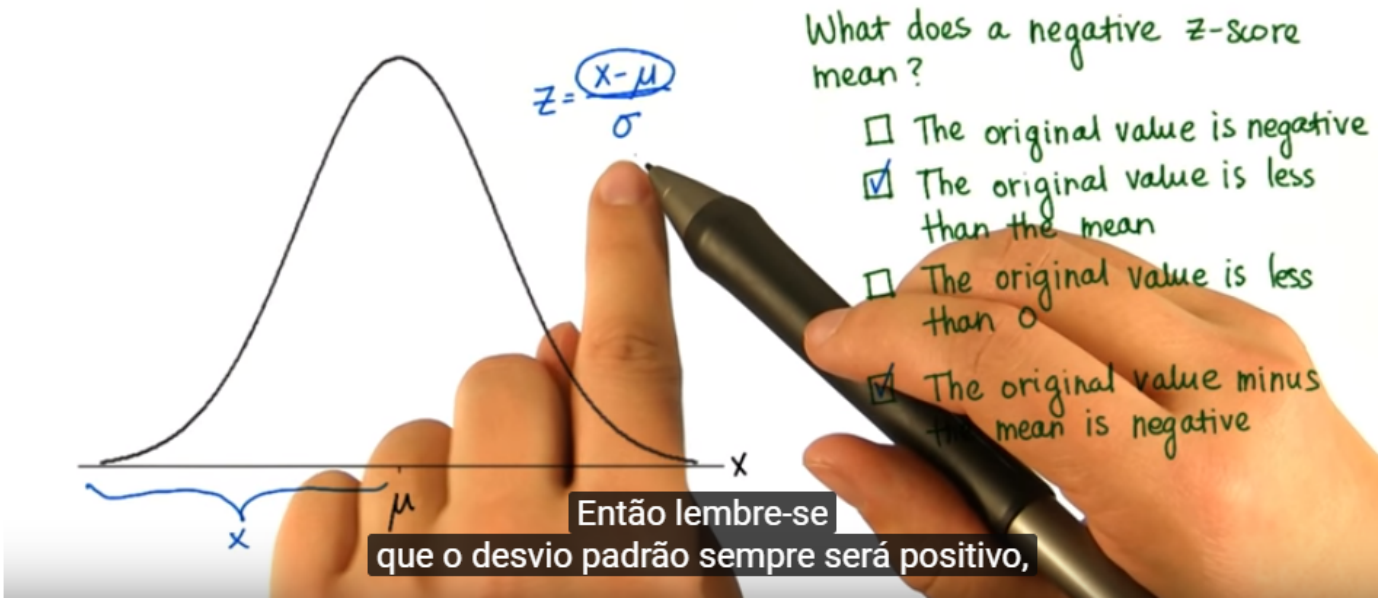
- ○ $\mu - x$
- ○ $x - \mu$
- ○ $(x-\mu)\sigma$
- ○ $\frac{x-\mu}{\sigma}$
- ○ $(\mu-x)\sigma$
- ✓ $\frac{(\mu-x)}{\sigma}$



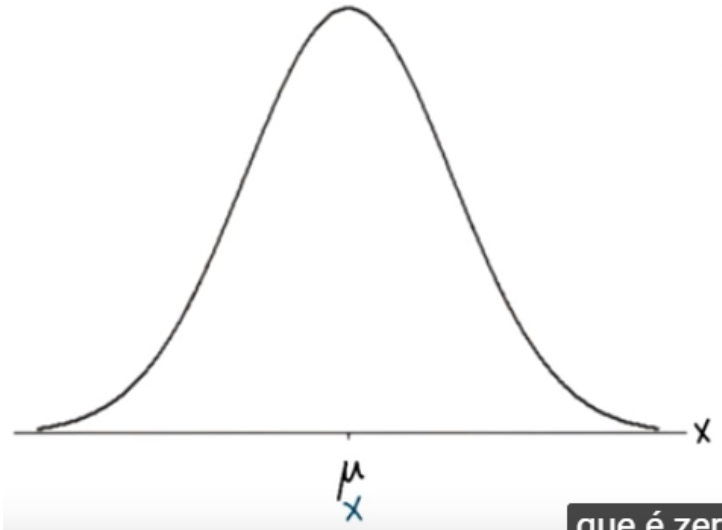
Então esta é a resposta correta para achar a distância em desvios padrão



Sempre será positivo porque é elevado ao quadrado



It's all from formula of Z. xi-x(mean). when x = x(mean) we get Z=0, so mean of Standardized distribution is 0. You can find in literature the phrase: "Distribution with mean = 0 and standard deviation = 1" So it's Standardized distribution



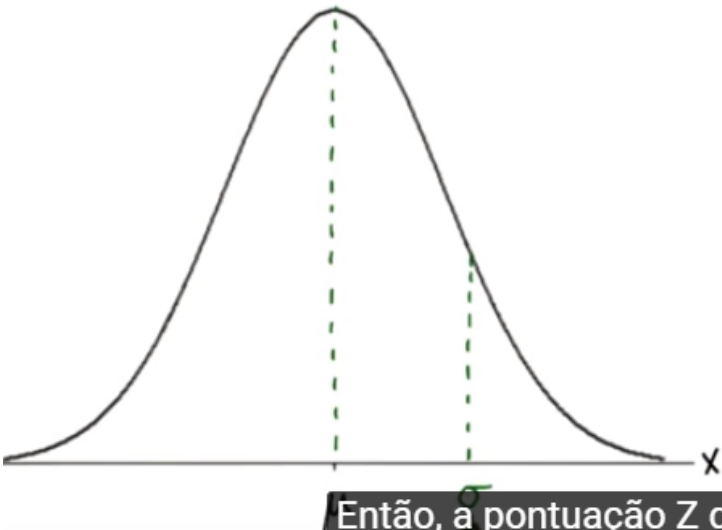
If we standardize a distribution by converting every value to a z-score, what will be the new mean of this standardized distribution?

$\mu = 0$

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{\mu - \mu}{\sigma} = \frac{0}{\sigma} = 0$$

que é zero.

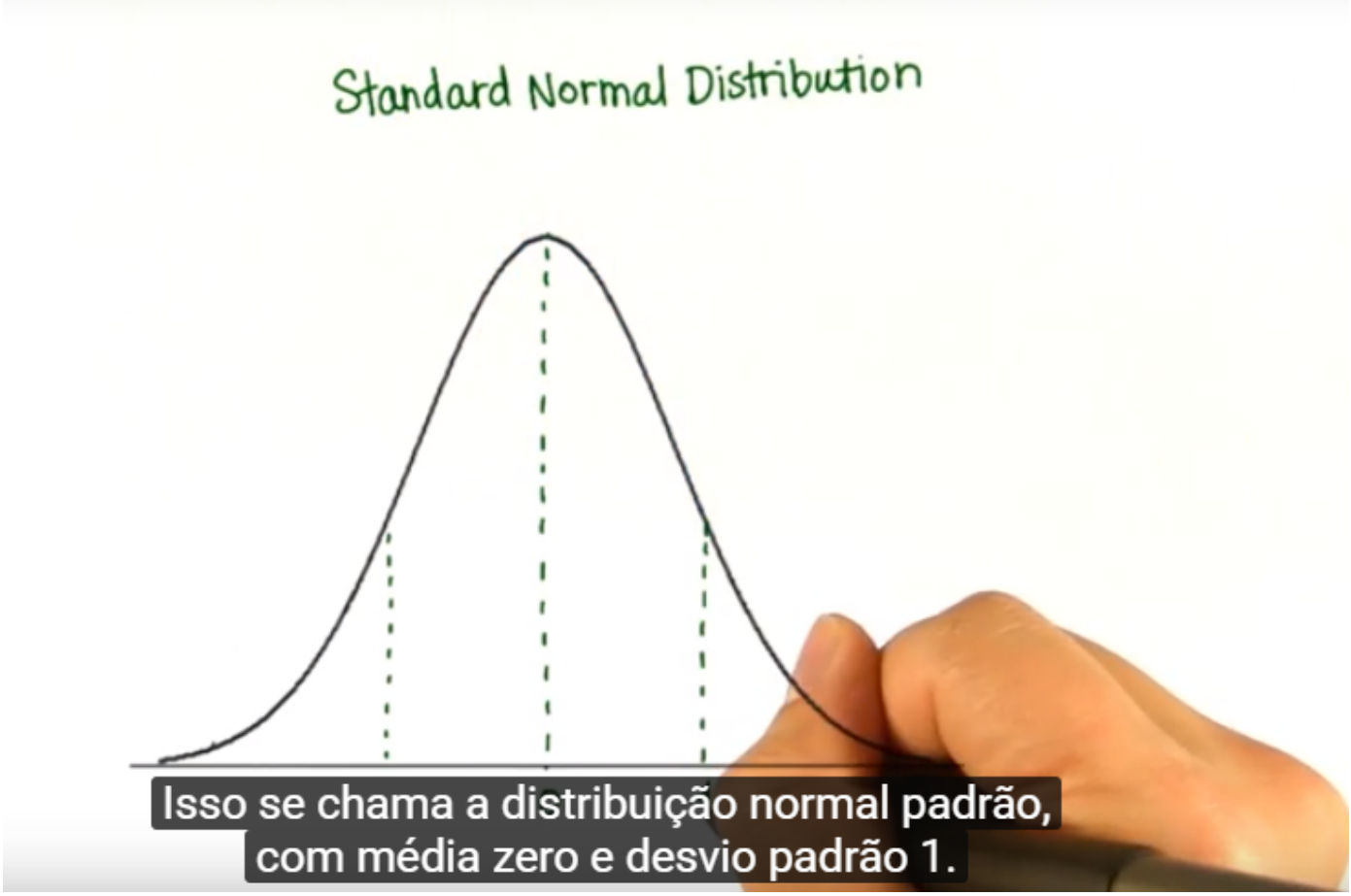


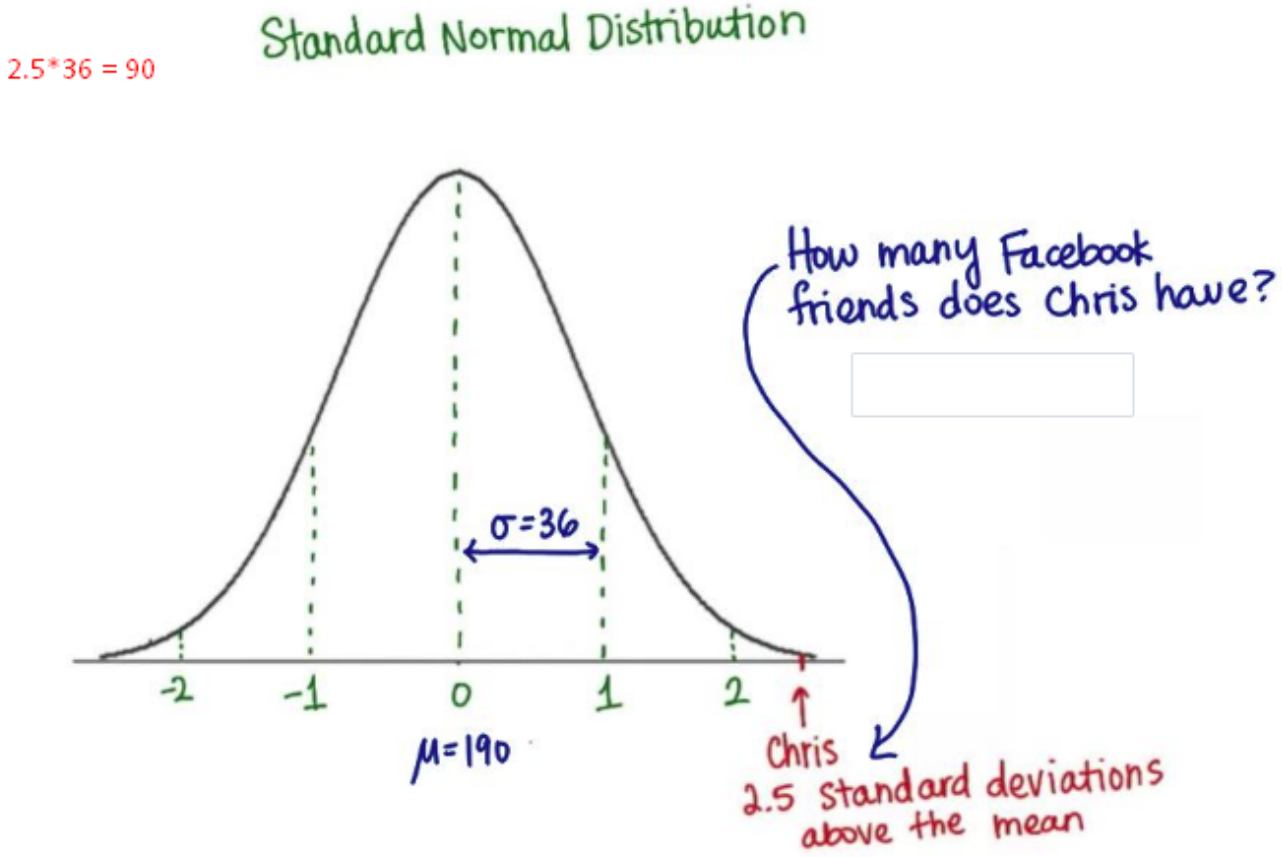
If we standardize a distribution by converting every value to a z-score, what will be the new standard deviation of this standardized distribution?

$$z = \frac{x - \mu}{\sigma}$$

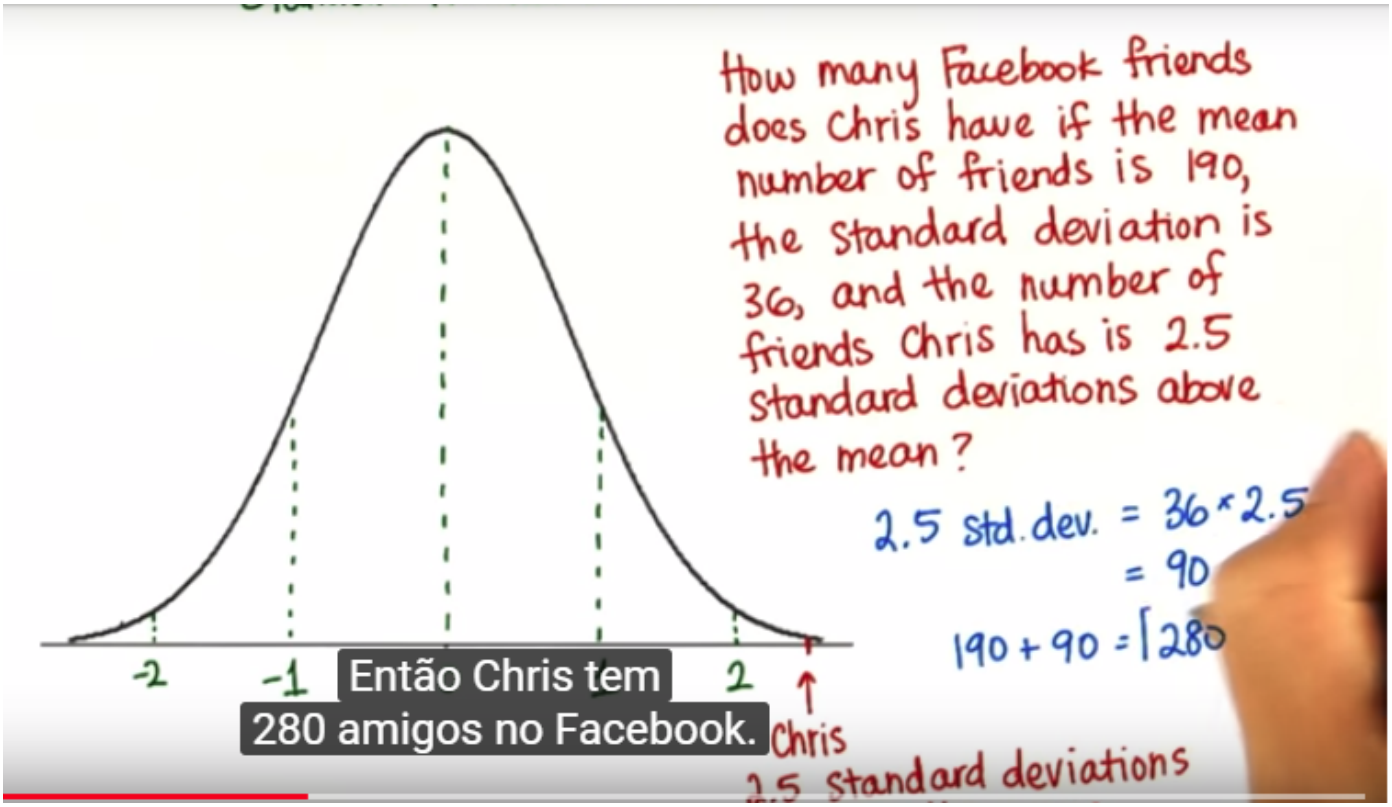
$$\frac{\sigma - 0}{\sigma} = \frac{\sigma}{\sigma} = 1$$

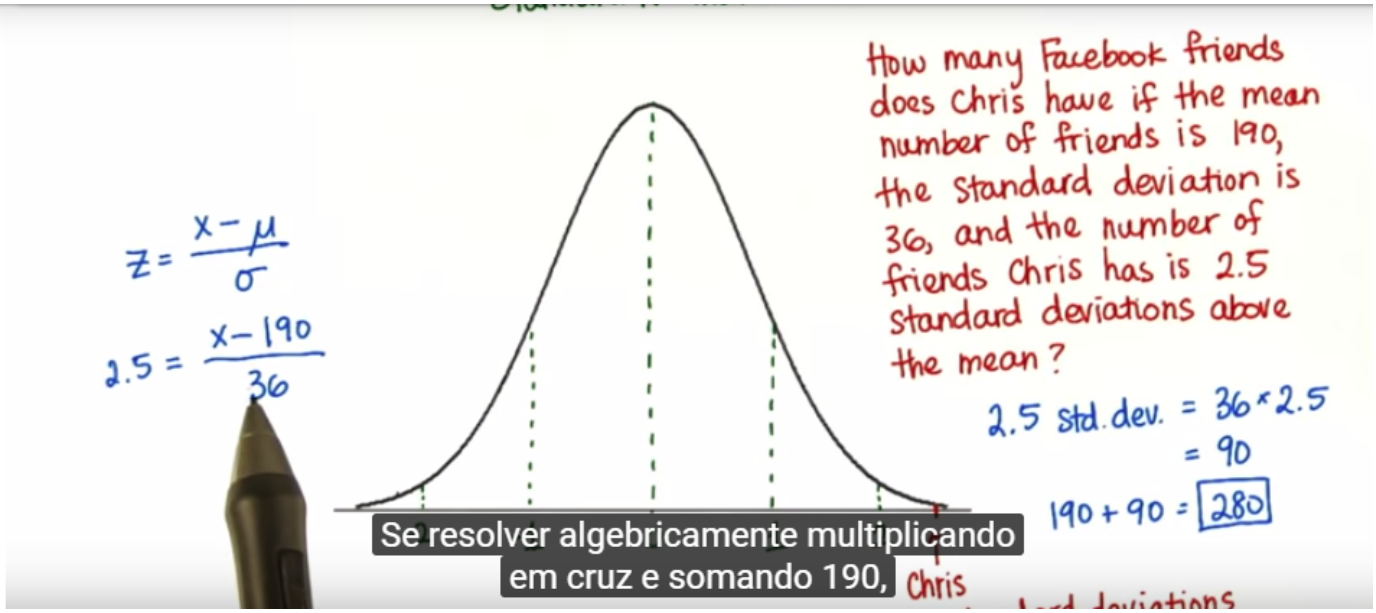
Então, a pontuação Z de qualquer valor que estiver a um desvio padrão da média

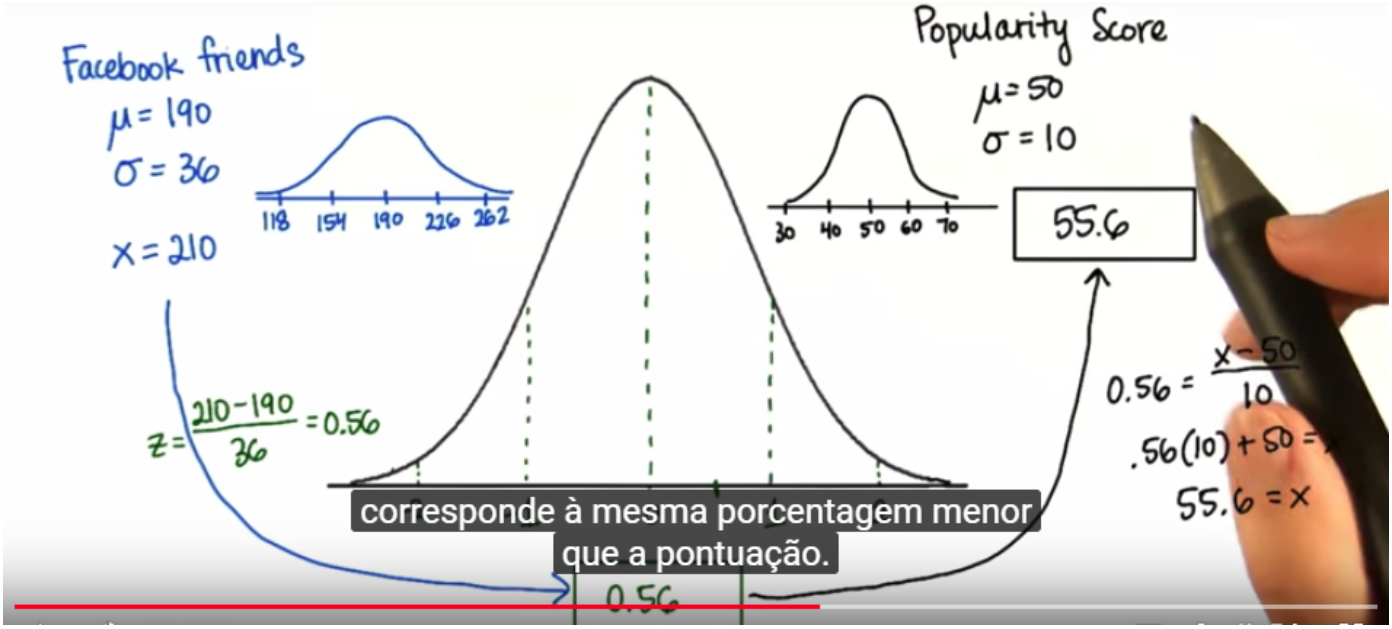




SOMA 190 + 90 = 280

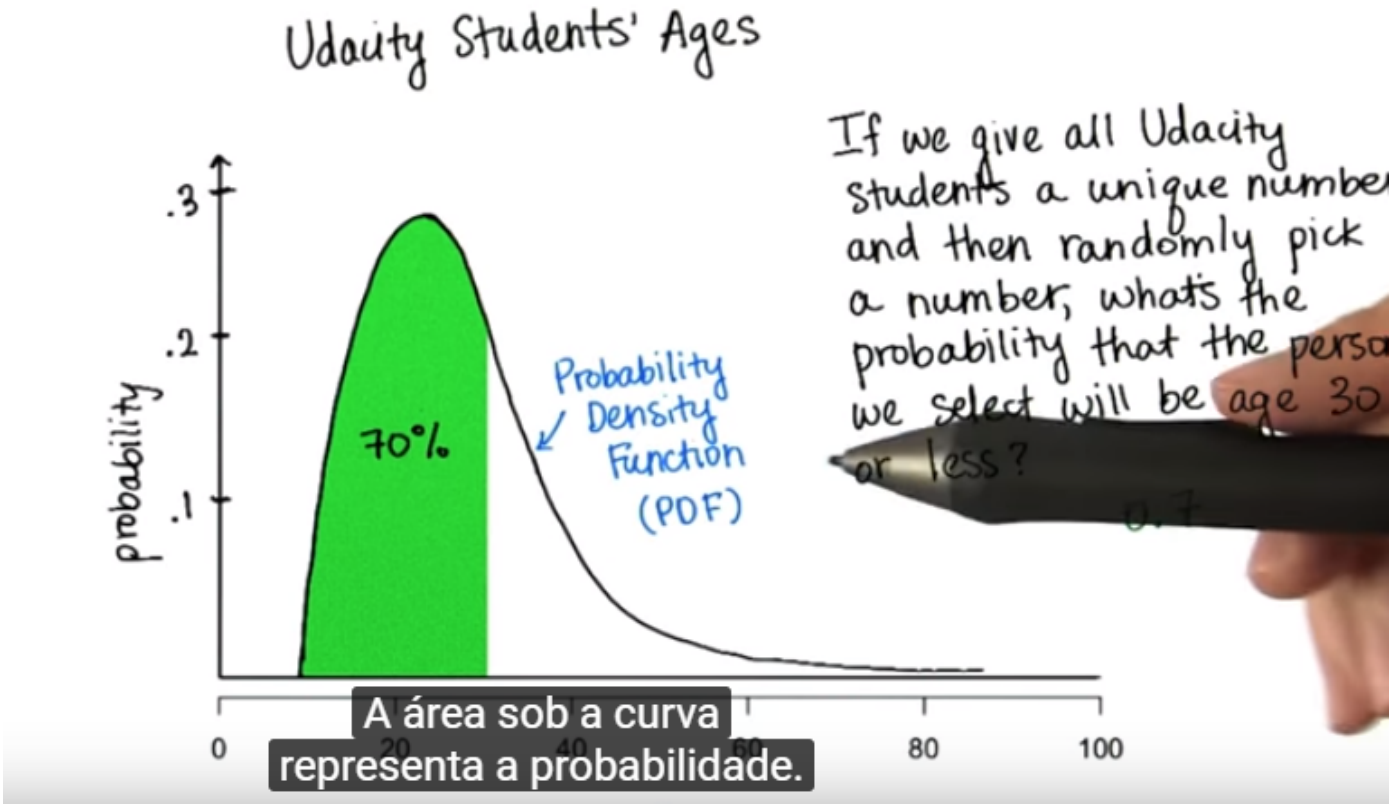

How many Facebook friends does Chris have if the mean number of friends is 190, the standard deviation is 36, and the number of friends Chris has is 2.5 standard deviations above the mean?
2.5 std. dev. = 36 × 2.5
= 90
190 + 90 = 280
-2
-1
2
Então Chris tem 280 amigos no Facebook.
Chris
2.5 standard deviations


How many Facebook friends does Chris have if the mean number of friends is 190, the standard deviation is 36, and the number of friends Chris has is 2.5 standard deviations above the mean?
$z = \frac{x - \mu}{\sigma}$
$2.5 = \frac{x - 190}{36}$
2.5 std. dev. = 36 × 2.5
= 90
190 + 90 = 280
Se resolver algebricamente multiplicando em cruz e somando 190,
Chris


Facebook friends
μ = 190
σ = 36
x = 210
118 154 190 226 262
Popularity Score
μ = 50
σ = 10
30 40 50 60 70
55.6
z = (210 − 190)/36 = 0.56
0.56 = (x − 50)/10
.56(10) + 50 = x
55.6 = x
corresponde à mesma porcentagem menor
que a pontuação.
0.56

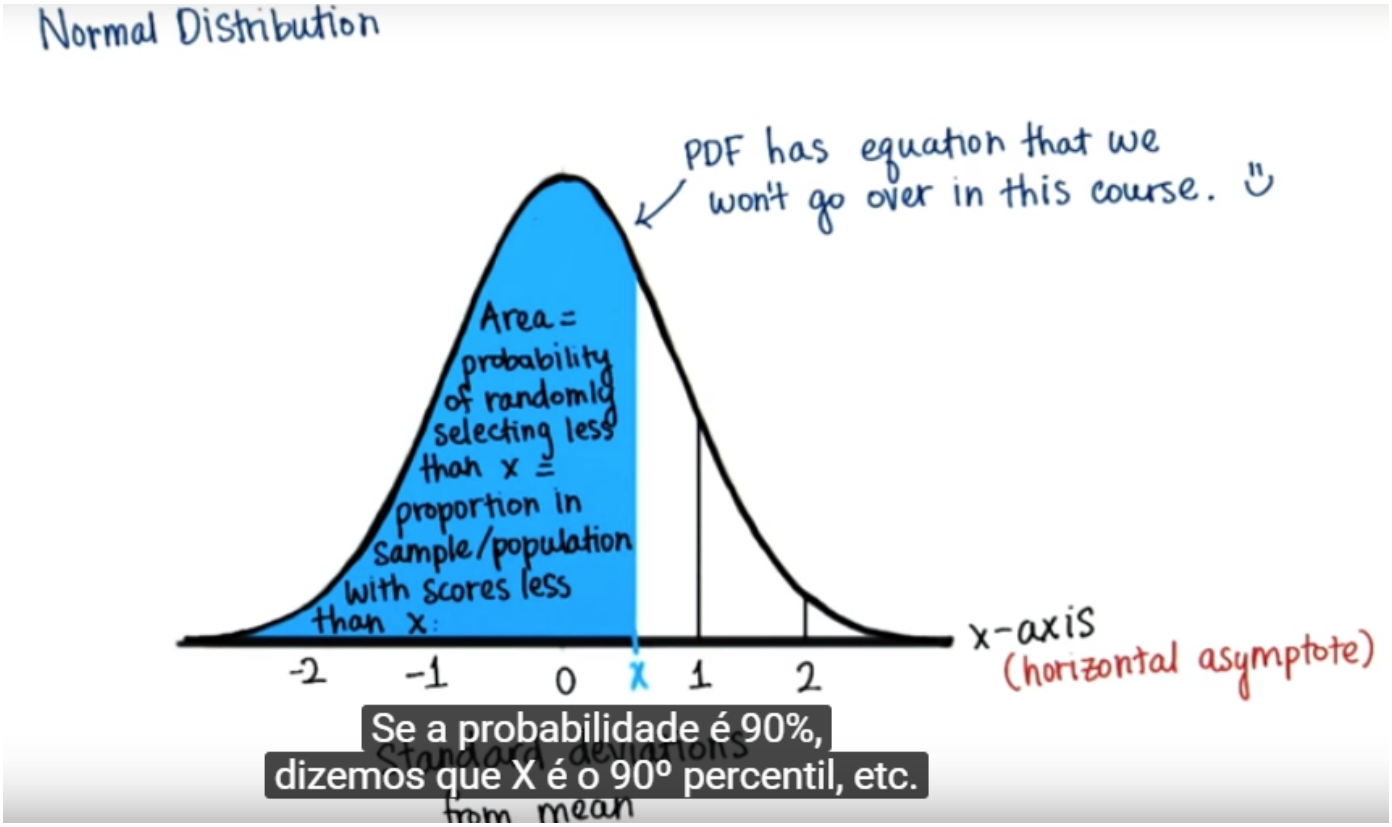
PDF


Normal Distribution
Probability Density Function
(PDF)
x-axis
-2 -1 0 1 2
Essa curva se chama função
e densidade de probabilidade,
from mean

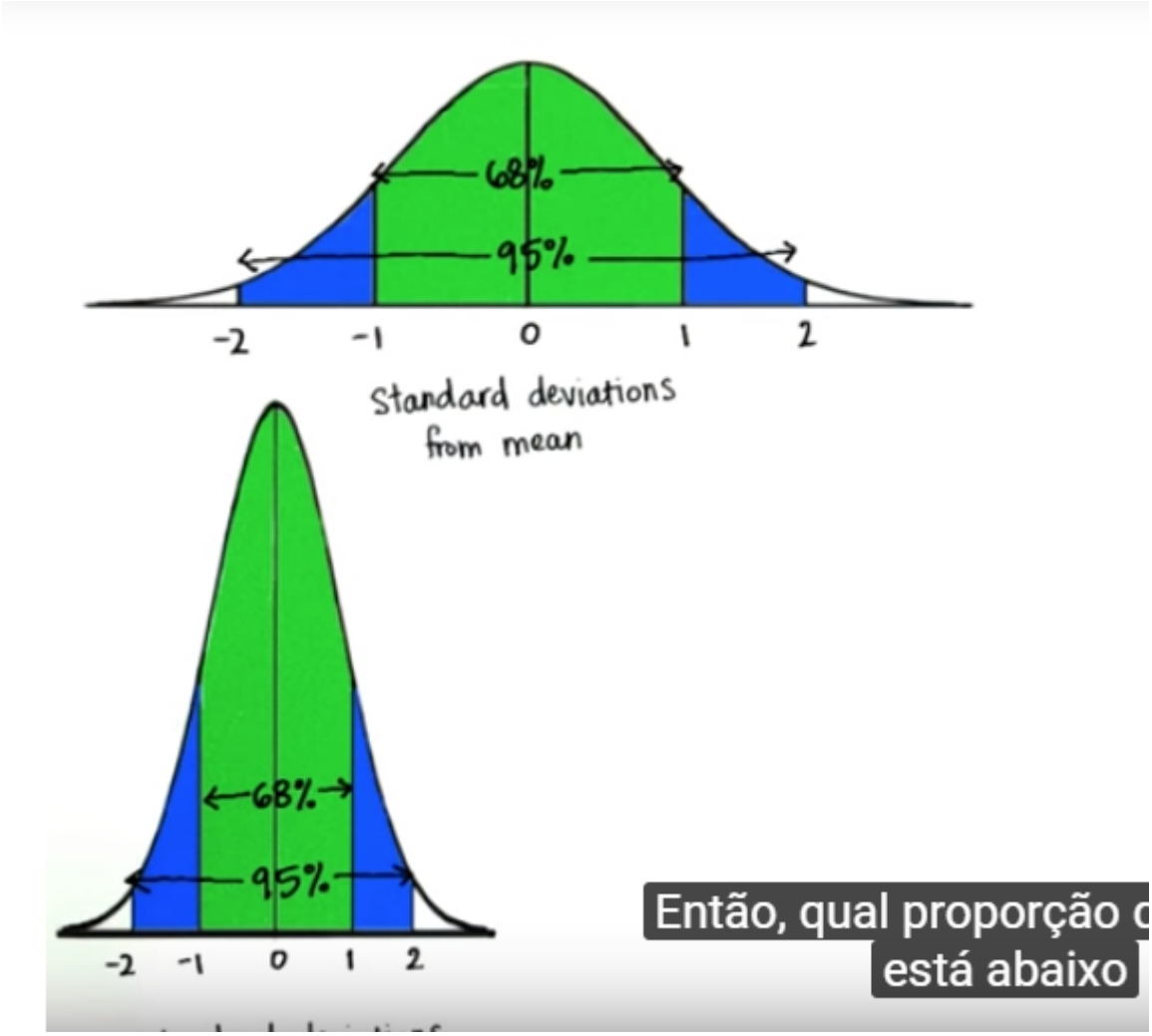


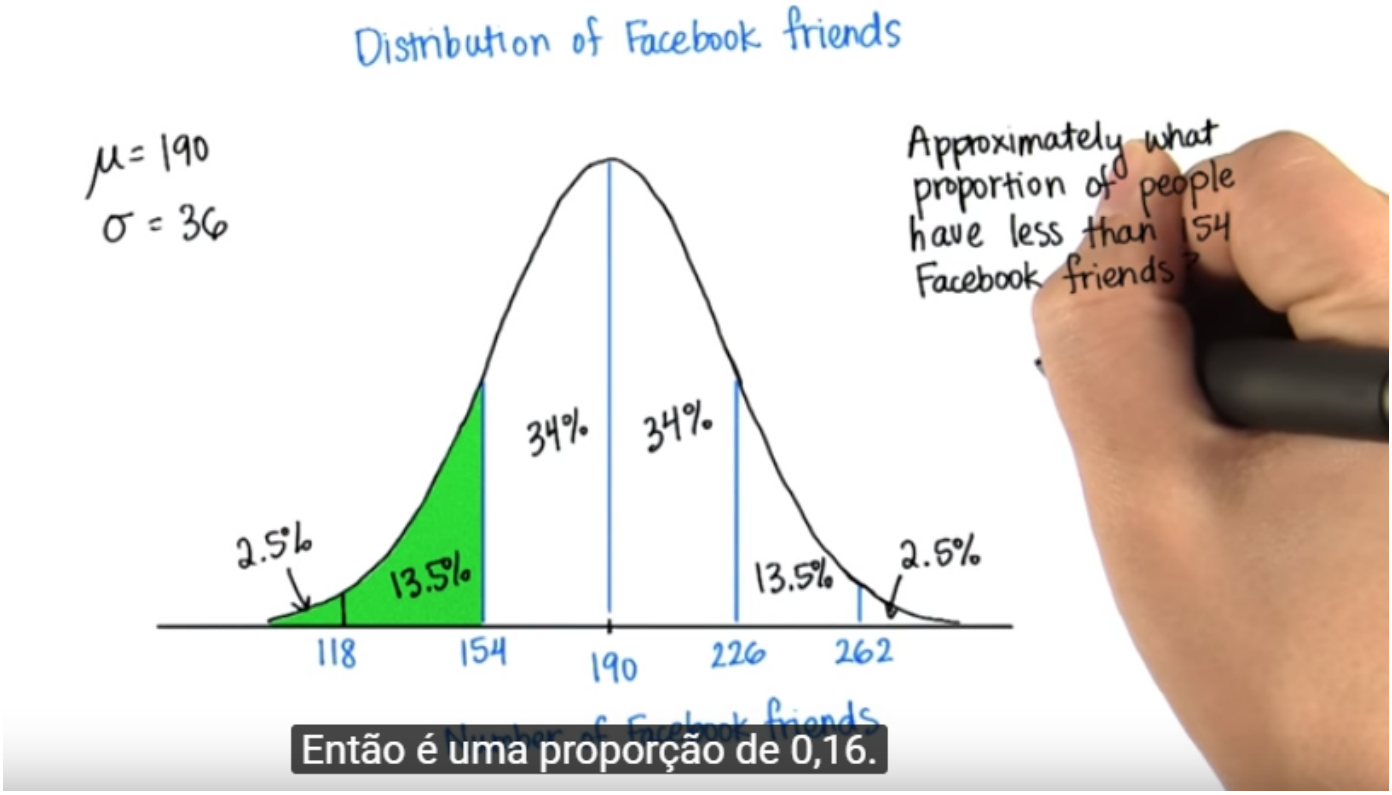
PDF regras

- As extremidades nunca tocam o eixo X (Consideramos isso, pois não podemos ter 100% de algo)
- Se existe um certo valor (x) a área sob a curva do negativo até X




-2 -1 0 1 2
Standard deviations from mean
-2 -1 0 1 2
Standard deviations from mean
e para distribuições normais, cerca de 68%
fica em um centro do desvio da média


68%
95%
-2 -1 0 1 2
Standard deviations from mean
68%
95%
-2 -1 0 1 2
Então, qual proporção d
está abaixo


Distribution of Facebook friends
μ = 190
σ = 36
Approximately what proportion of people have less than 154 Facebook friends?
34%
34%
2.5%
13.5%
13.5%
2.5%
118
154
190
226
262
Então é uma proporção de 0,16.

In [ ]: