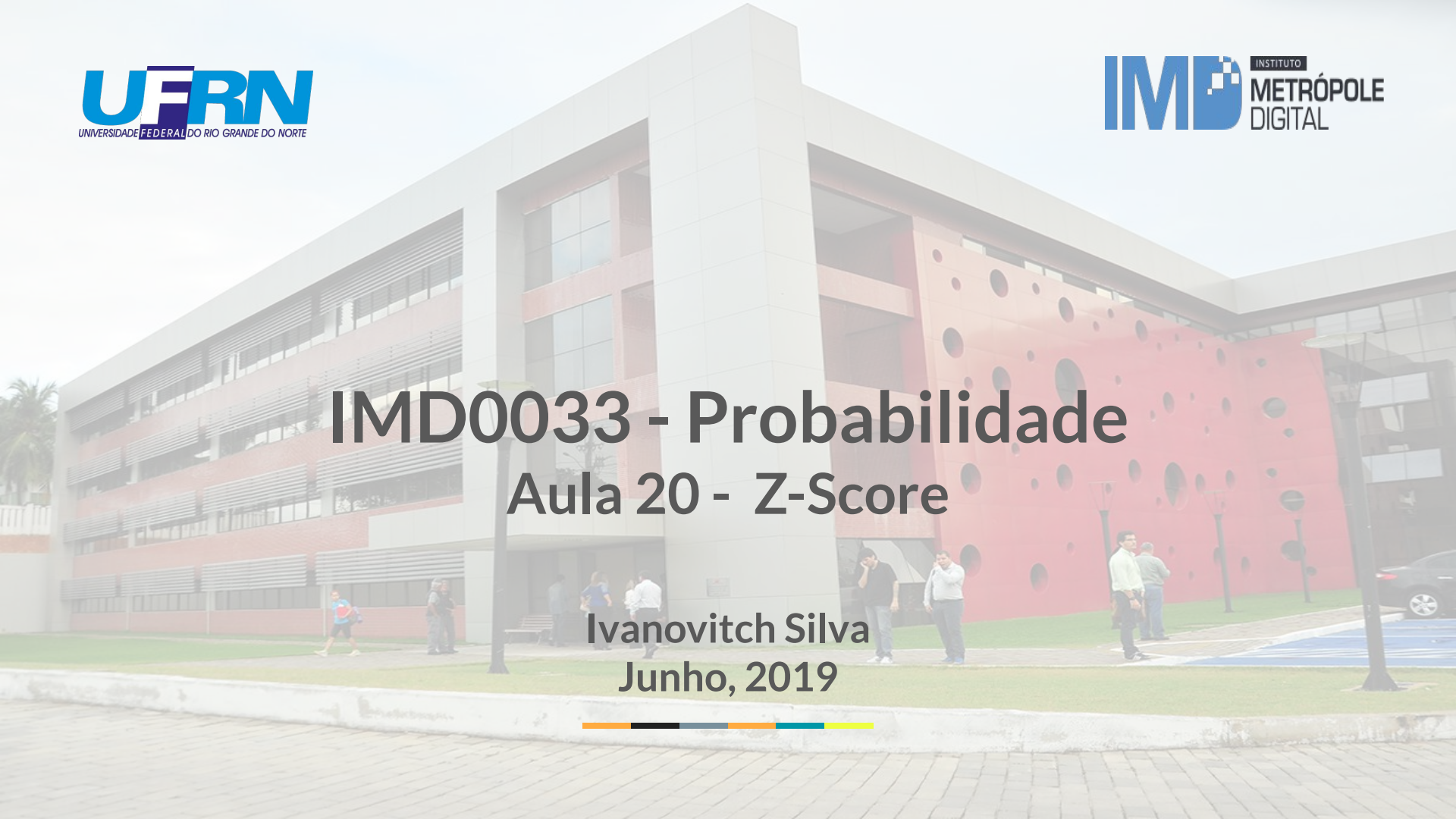


IMD0033 - Probabilidade

Aula 20 - Z-Score

Ivanovitch Silva

Junho, 2019

Agenda

- Definição de Z-Score
- Distribuição padrão
- Entendendo melhor os pontos fora da curva
- Z-Score como medida de comparação
- Z-Table
- Transformação de Z-Score em valor

Atualizar o repositório

```
git clone https://github.com/ivanovitchm/imd0033_2019_1.git
```

Ou ....

```
git pull
```

Problem: house prices

	Order	PID	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Pool QC	Yr Sold	Sale Type	Sale Condition	SalePrice
0	1	526301100	20	RL	141.0	131770	Pave	0	5	2010	WD	Normal	215000
1	2	526350040	20	RH	80.0	11622	Pave	0	6	2010	WD	Normal	105000
2	3	526351010	20	RL	81.0	14267	Pave	0	6	2010	WD	Normal	172000
3	4	526353030	20	RL	93.0	11160	Pave	0	4	2010	WD	Normal	244000
4	5	527105010	60	RL	74.0	13830	Pave	0	3	2010	WD	Normal	189900

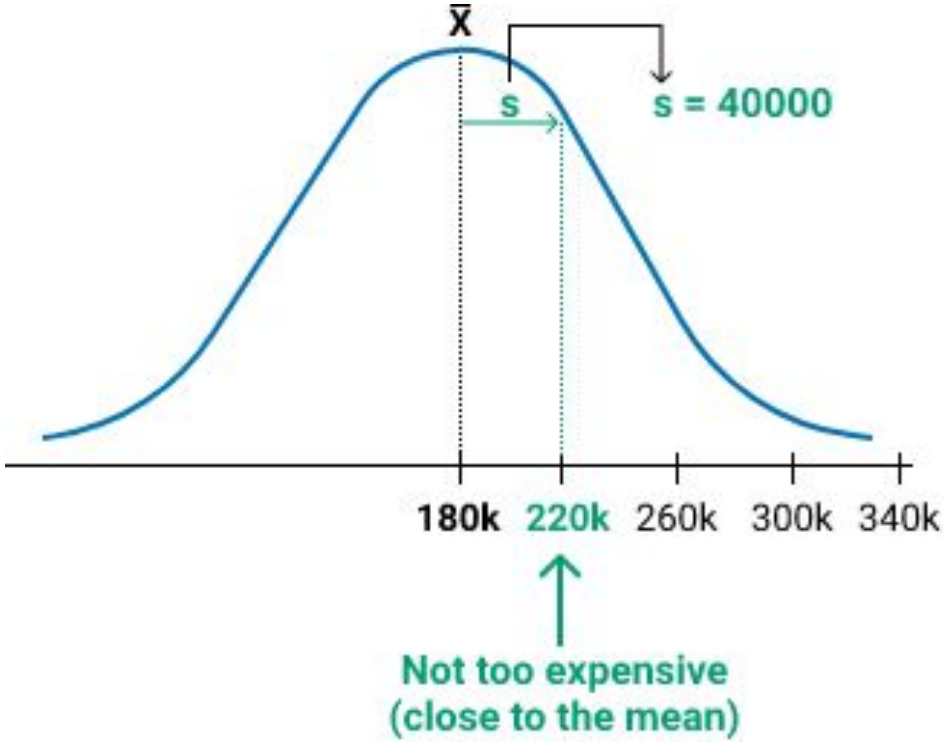
Is a house costing US 220,000 cheap, expensive, or average-priced?

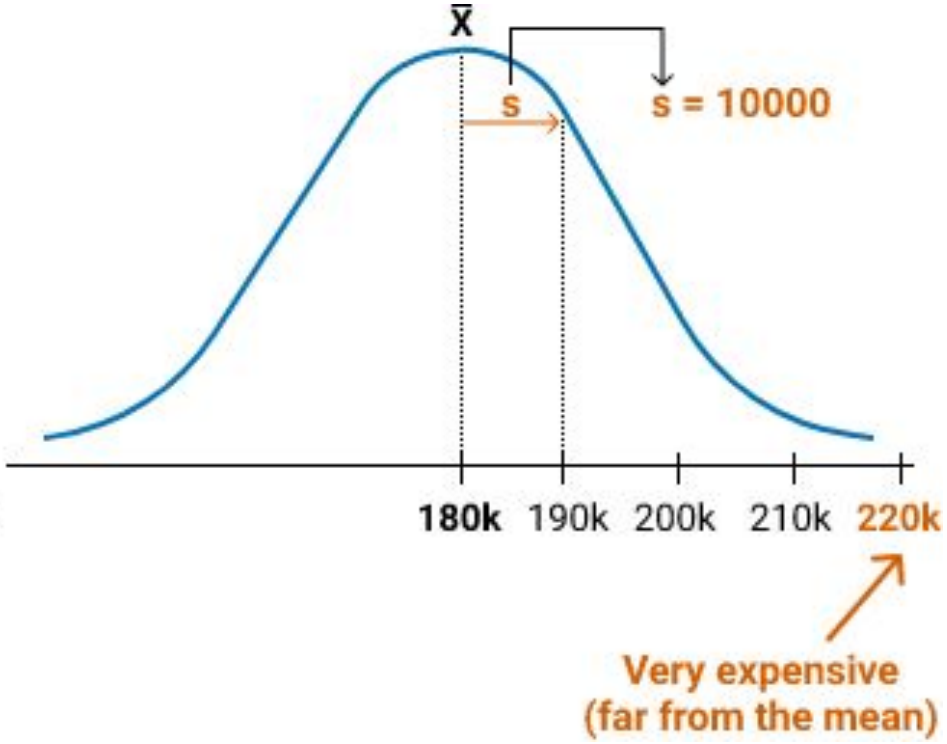
```
print(houses['SalePrice'].mean())
180796.0600682594
```

The sampled house (US 220,000) is clearly more expensive than the average house (roughly US180,796)!!!

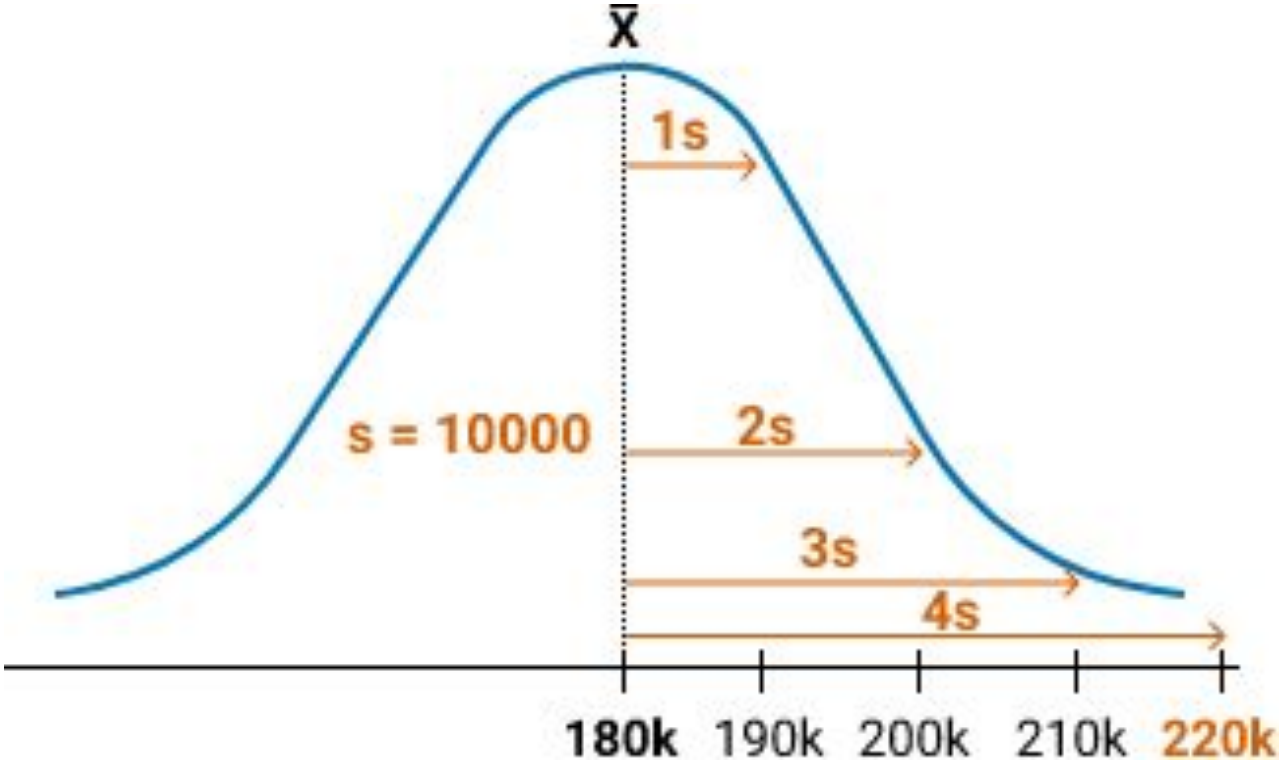
But is this price **slightly above** the average or **extremely** above the average?

The answer depends on the **standard deviation** of the distribution of sale prices.

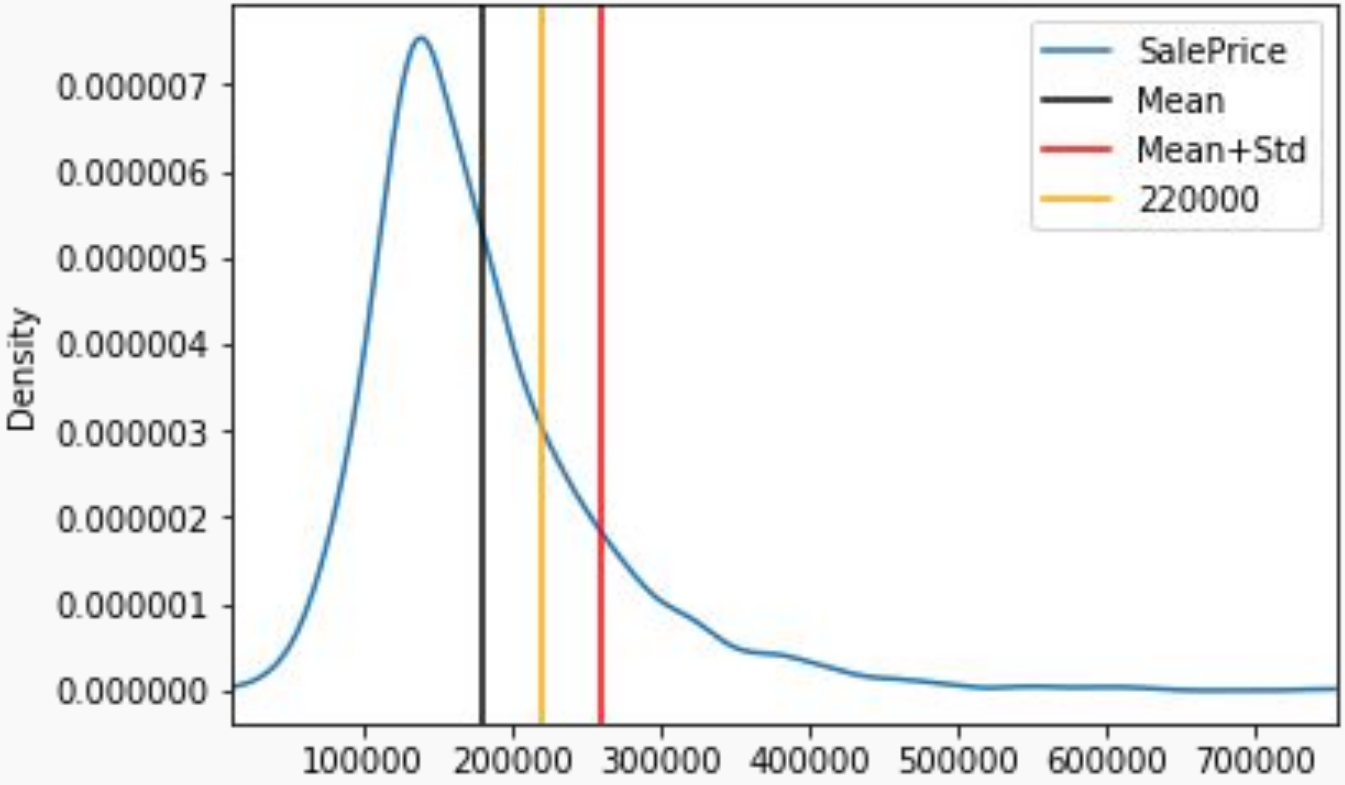

X̄
s
s = 40000
180k
220k
260k
300k
340k
Not too expensive
(close to the mean)


X̄
s
s = 10000
180k
190k
200k
210k
220k
Very expensive
(far from the mean)

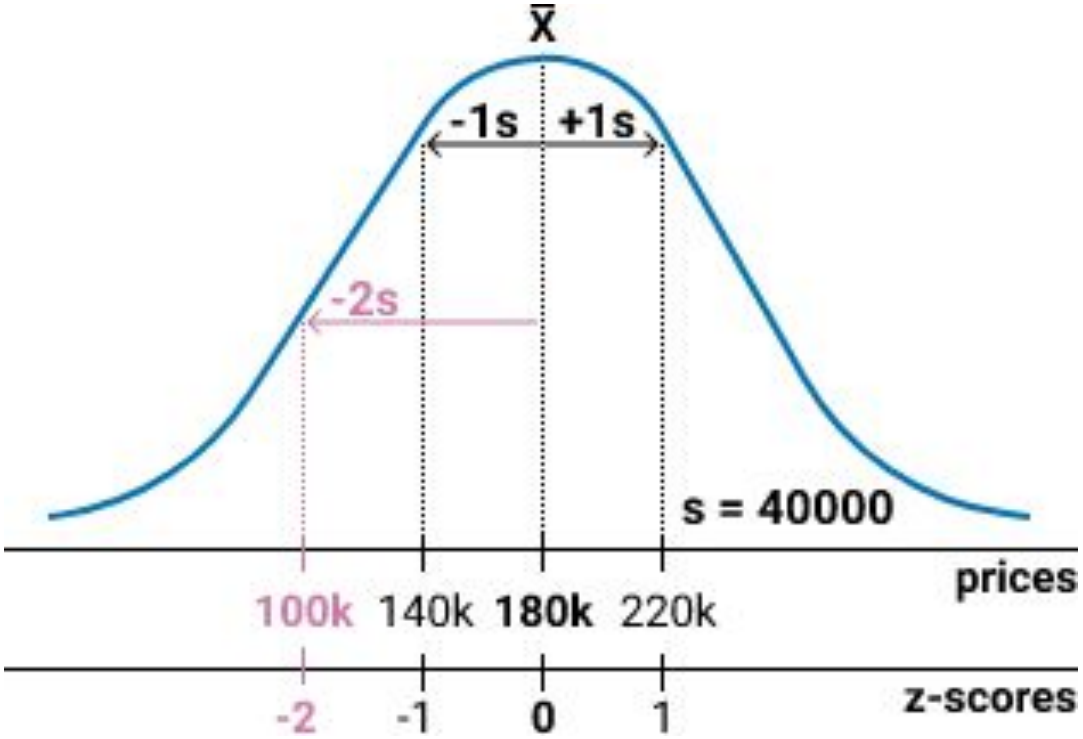
Number of Standard Deviations



Examine the graph and figure out whether a price of US 220,000 is very expensive.



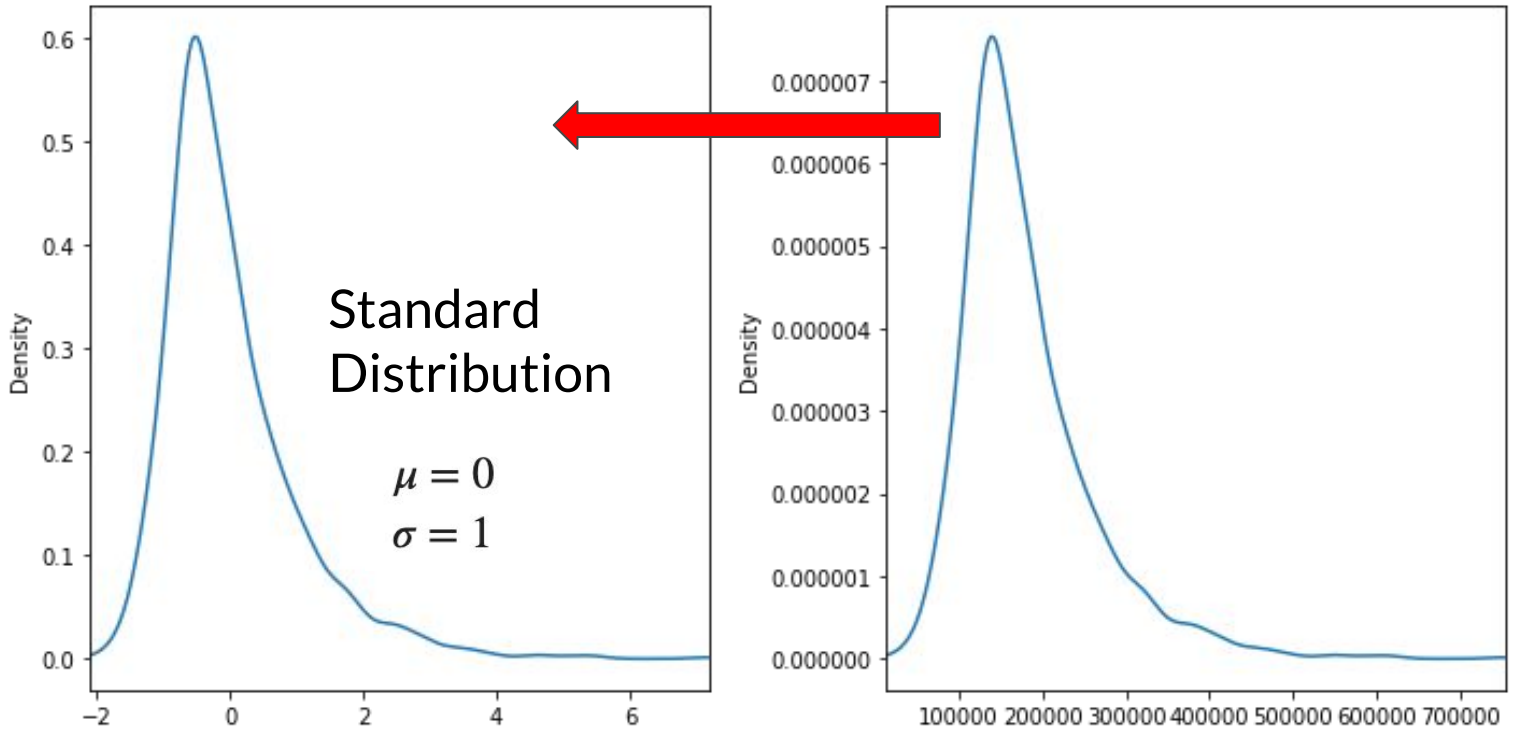
Z-Score

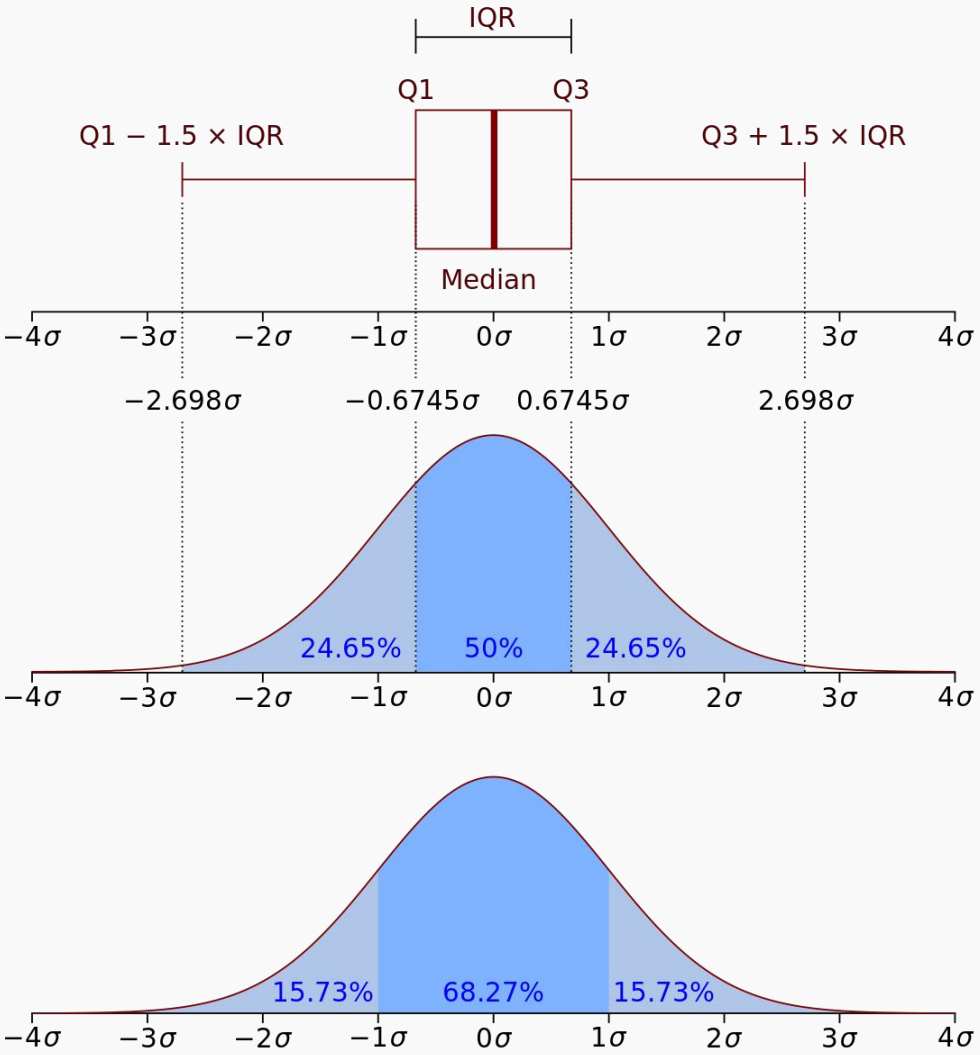


$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{x - \bar{x}}{s}$$

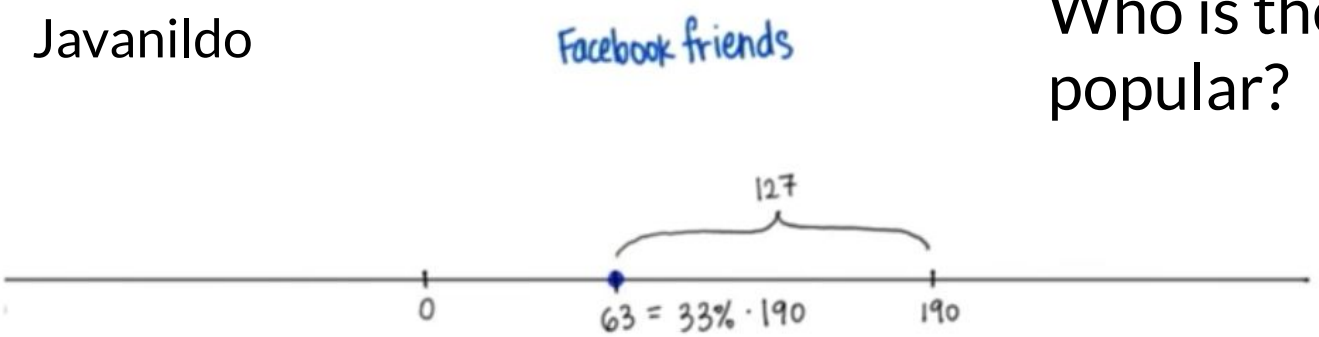
Transforming Distribution




IQR
Q1
Q3
Q1 − 1.5 × IQR
Q3 + 1.5 × IQR
Median
−4σ
−3σ
−2σ
−1σ
0σ
1σ
2σ
3σ
4σ
−2.698σ
−0.6745σ
0.6745σ
2.698σ
24.65%
50%
24.65%
15.73%
68.27%
15.73%

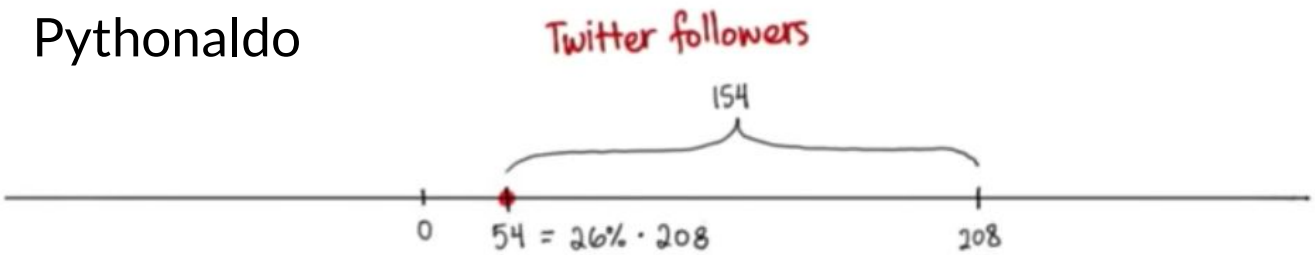
Using Standardization for Comparisons

Javanildo

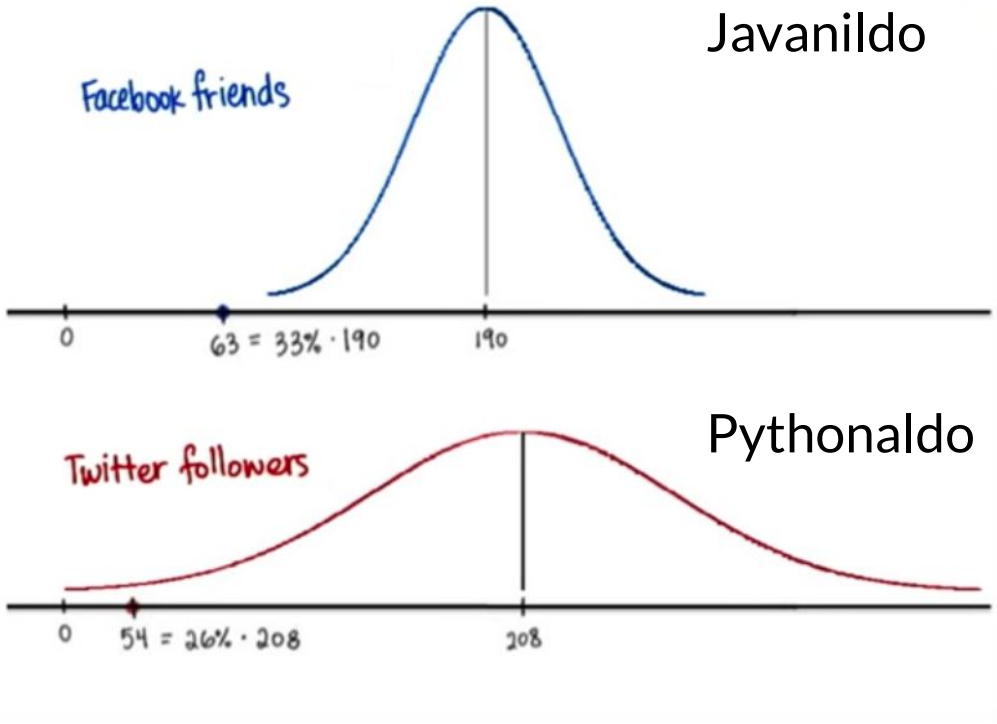


Who is the most popular?

Pythonaldo



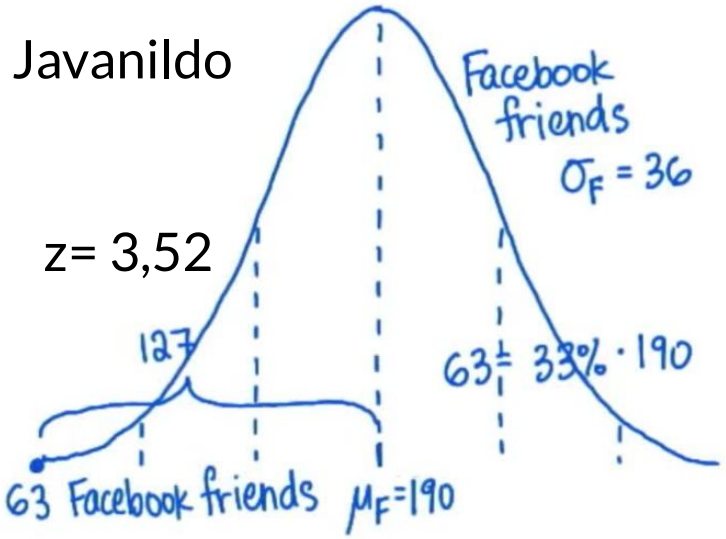
Quiz: who is the most popular?



Quiz: who is the most popular?

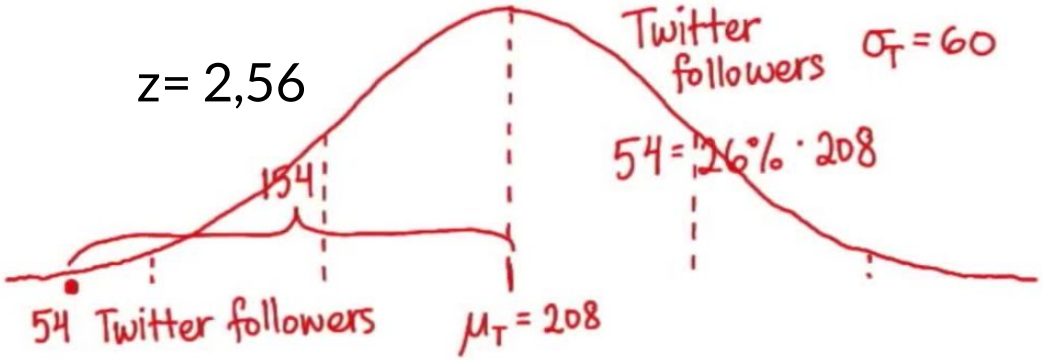
Javanildo

z= 3,52

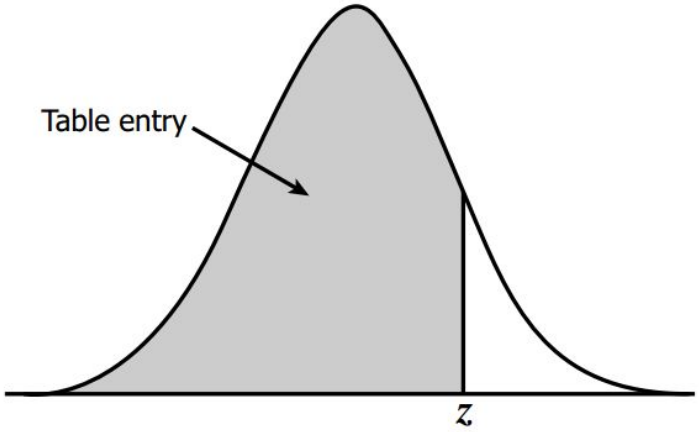


Pythonaldo

z= 2,56



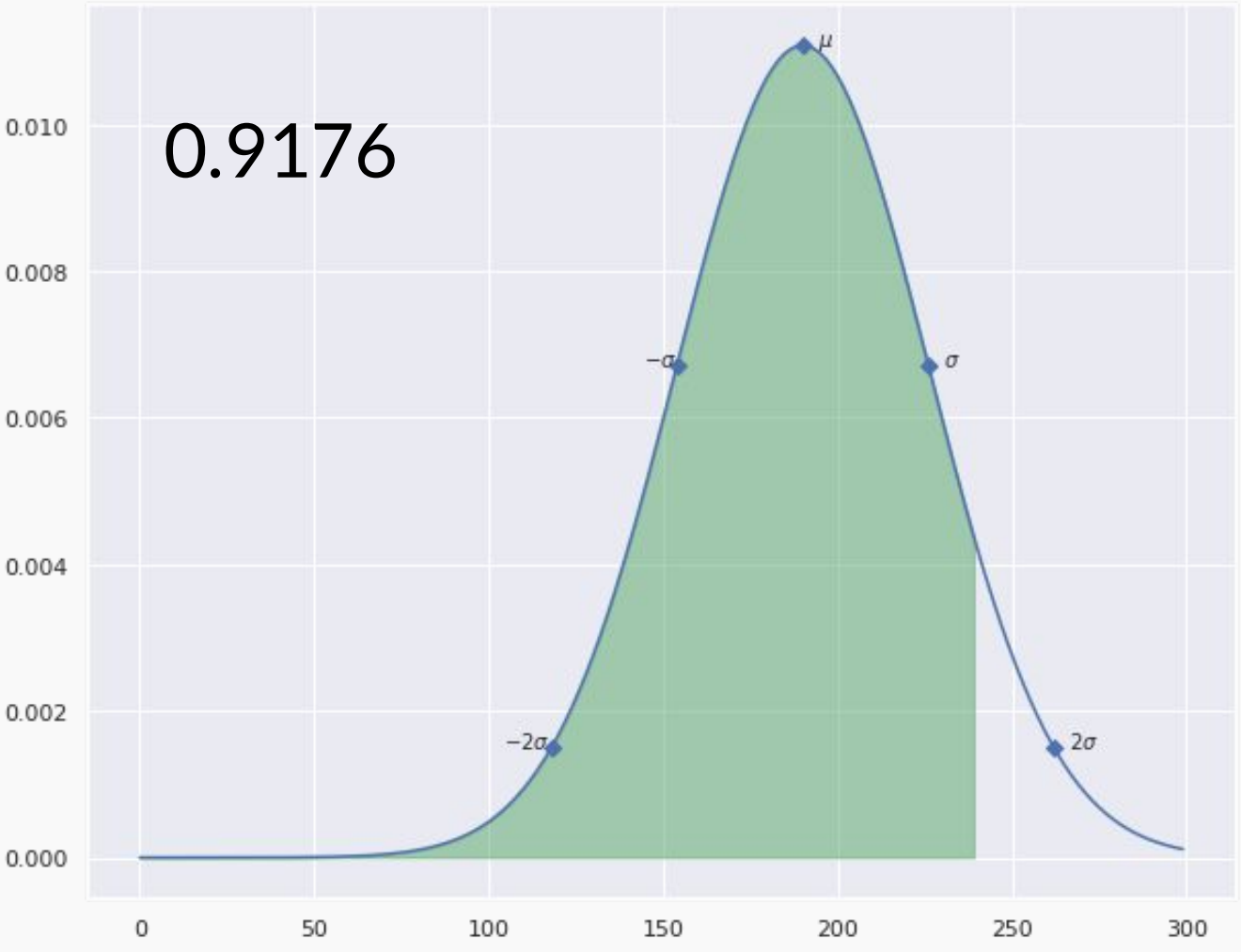
Z-Table



http://www.z-table.com/

Facebook example:
μ = 190
σ = 36
Xi = 240

What is the percentage of people who have less than 240 facebook friends?


0.9176
0.010
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
250
300
μ
−σ
σ
−2σ
2σ

Converting back from Z-Score

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

$$x = \sigma z + \mu$$


MacBook Pro