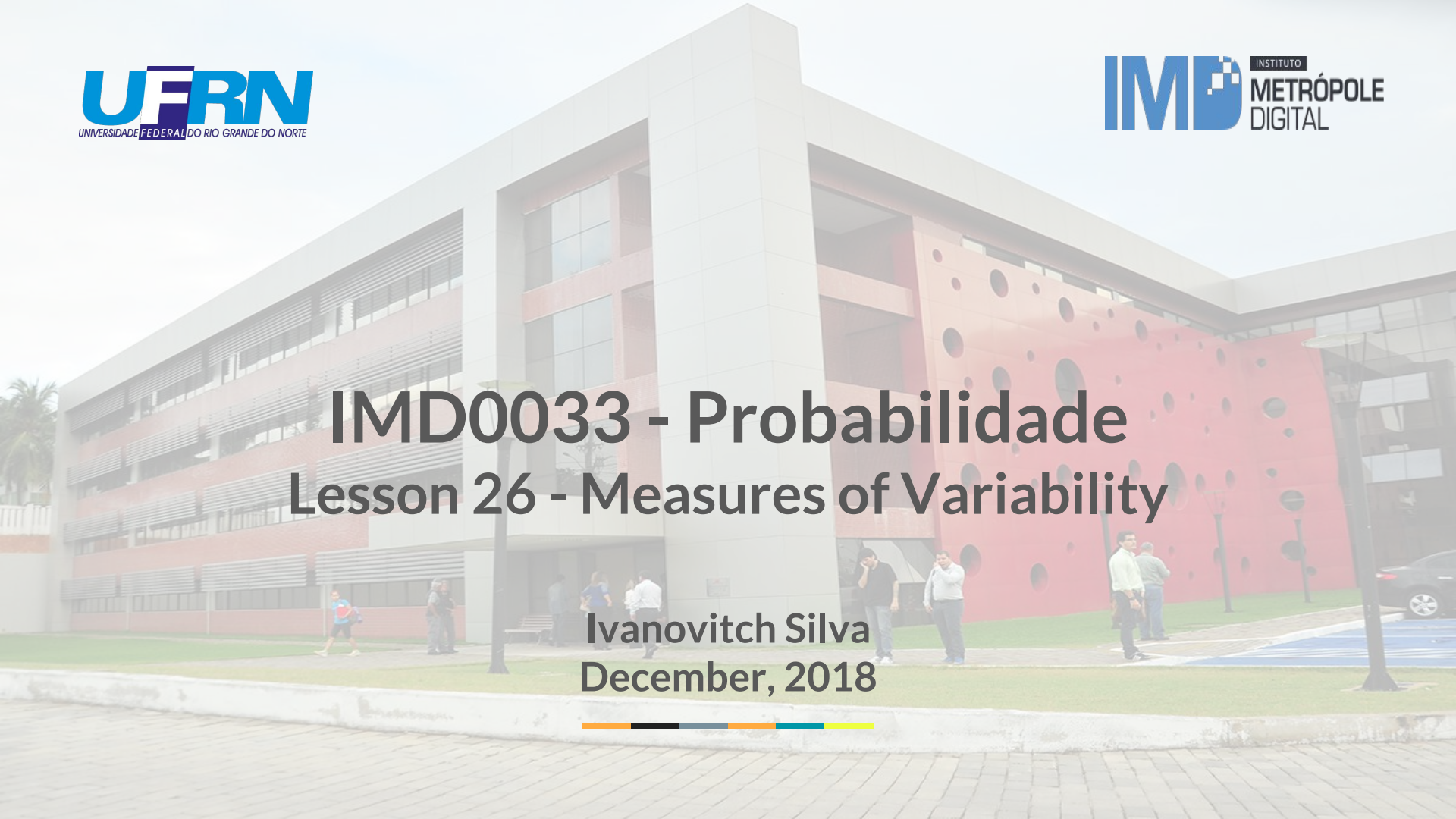


IMD0033 - Probabilidade

Lesson 26 - Measures of Variability

Ivanovitch Silva

December, 2018

Agenda

- The range
- Average distance
- Mean absolute deviation
- Variance and standard deviation
- The sample standard deviation
- Bessel's correction
- Z-Score

Atualizar o repositório

```
git clone https://github.com/ivanovitchm/imd0033_2018_2.git
```

Ou ....

```
git pull
```

The range

So far we've focused entirely on summarizing distributions using the **mean**, the **weighted mean**, the **median**, and the **mode**. An interesting distribution property we haven't yet discussed is **variability**.

The values of the distribution A don't vary

$$A = [4, 4, 4, 4]$$

$$B = [0, 8, 0, 8]$$

What variability value should we assign to distribution B?

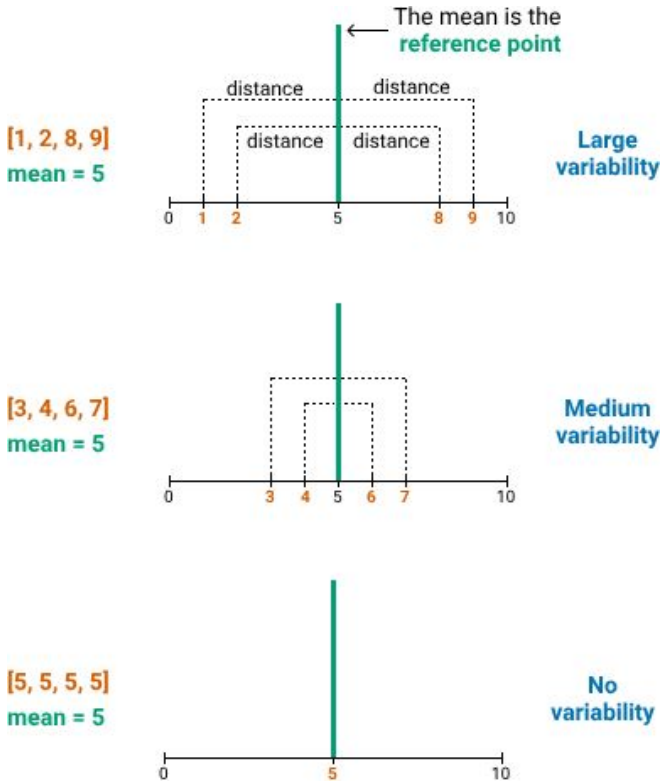
The range

$$A = [4, 4, 4, 4] \qquad max(A) - min(A) = 4 - 4 = 0$$

$$B = [0, 8, 0, 8] \qquad max(B) - min(B) = 8 - 0 = 8$$

$$range(X) = max(X) - min(X)$$

The average distance



The problem with the **range** is that it considers only two values in the distribution — the **minimum** and the **maximum** value.

$$C = [1, 1, 1, 1, 1, 1, 1, 1, 1, 21]$$

$$max(C) - min(C) = 21 - 1 = 20$$

$$average\ distance = \frac{\overbrace{(x_1 - \mu)}^{distance} + \overbrace{(x_2 - \mu)}^{distance} + \ldots + \overbrace{(x_N - \mu)}^{distance}}{N} = \frac{\overbrace{\sum_{i=1}^{n} \overbrace{(x_i - \mu)}^{distance}}^{total\ distance}}{N}$$

Mean absolute deviation

Values that are **below** the mean

$x_i - \mu$	Distance
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2
1 - 3	- 2

Total = - 18

Values that are **above** the mean

$x_i - \mu$	Distance
21 - 3	+ 18

Total = + 18

$$\text{mean absolute distance} = \frac{|x_1 - \mu| + |x_2 - \mu| + \ldots + |x_N - \mu|}{N} = \frac{\sum_{i=1}^{N} |x_i - \mu|}{N}$$

$$C = [1, 1, 1, 1, 1, 1, 1, 1, 1, 21]$$

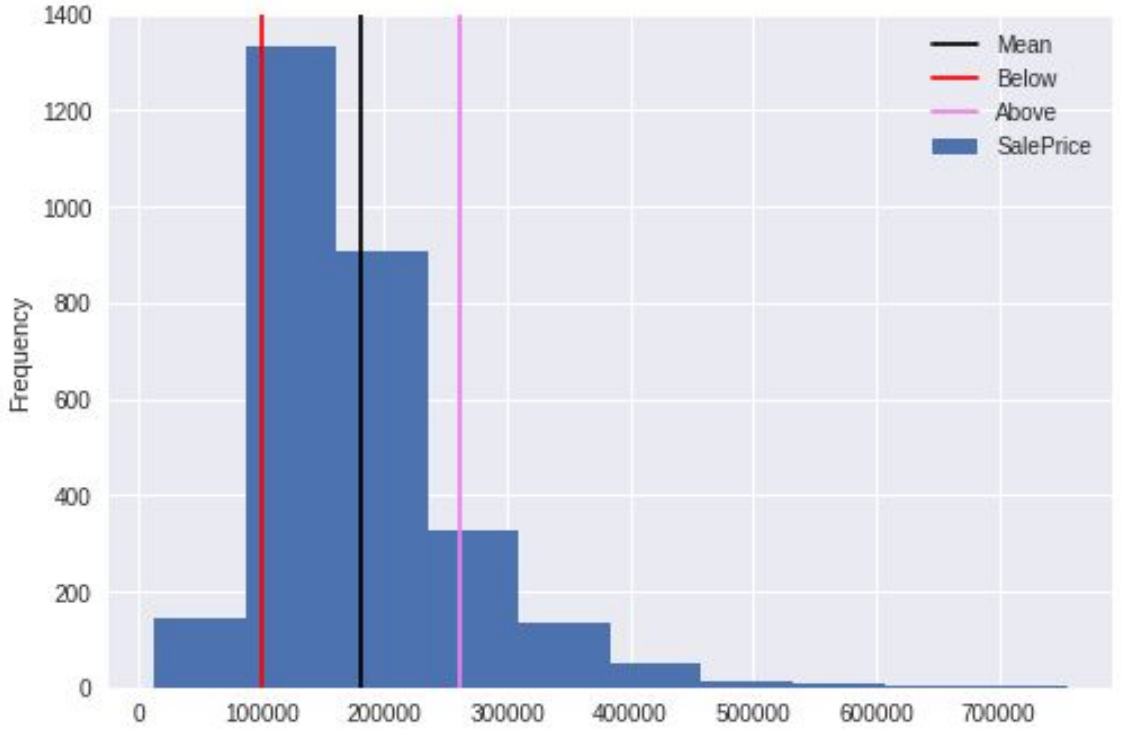
$$\text{average distance} = \frac{-18 + 18}{10} = \frac{0}{10}$$

Variance and Standard Deviation

$$\text{mean squared distance} = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_N - \mu)^2}{N} = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

$$\text{standard deviation} = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_N - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

Average variability around the mean

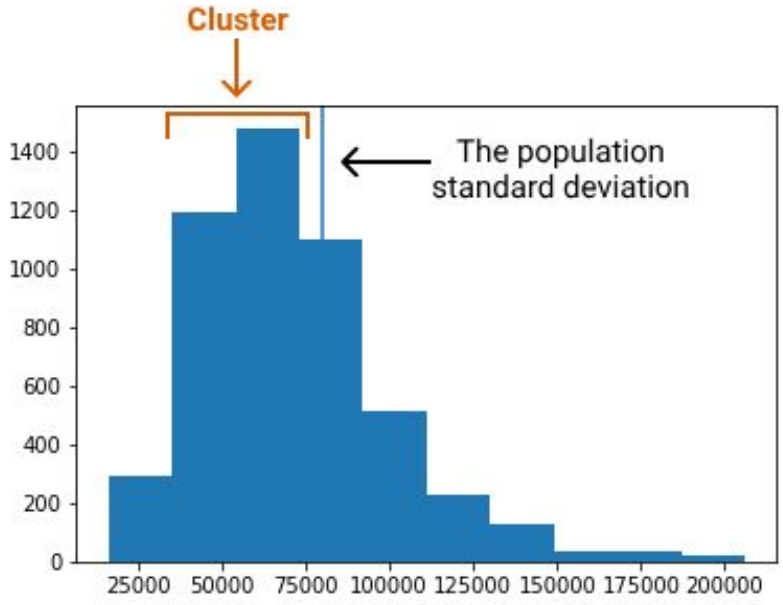


The sample standard deviation

$$SD = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \ldots + (x_N - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

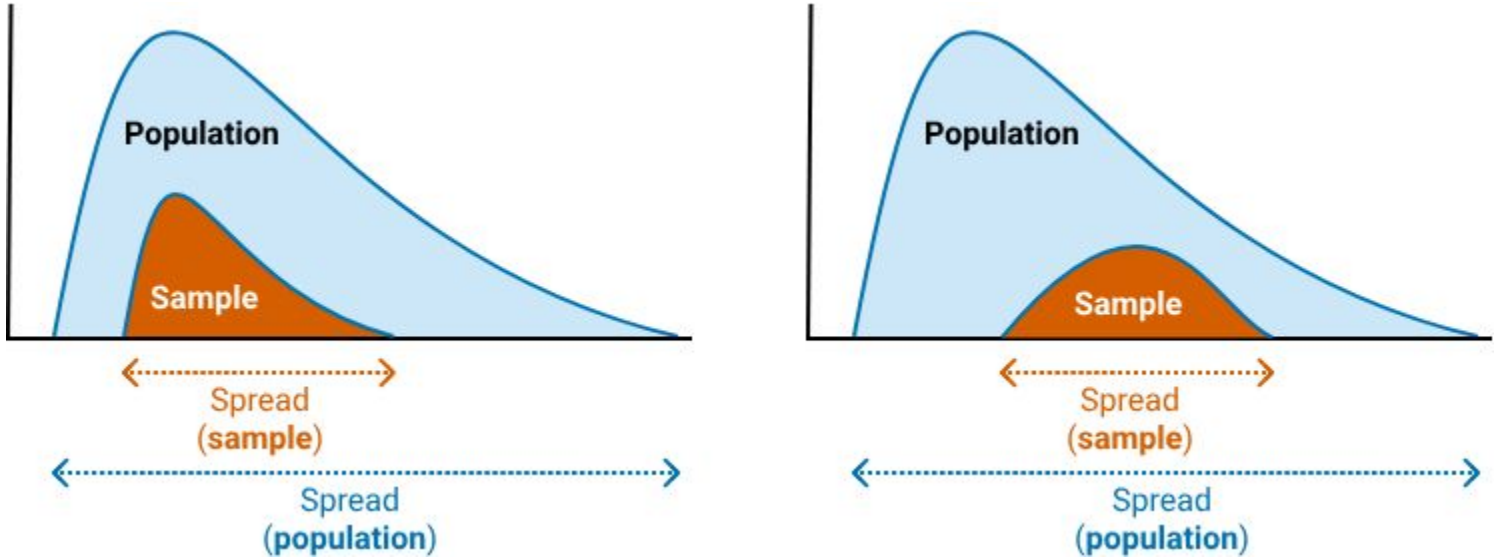
$$SD_{sample} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

The sample standard deviation



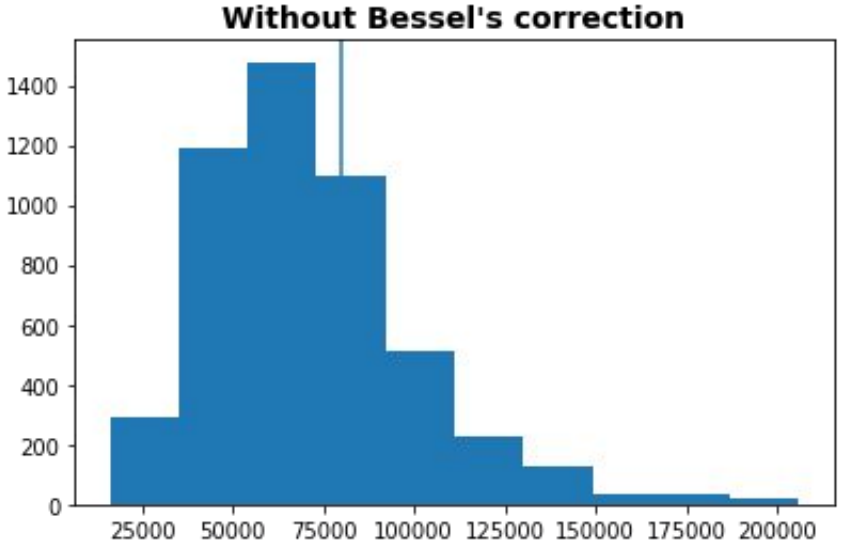

```
for i in range(5000):
    sample = houses['SalePrice'].sample(10, random_state = i)
    st_dev = standard_deviation(sample)
    st_devs.append(st_dev)

plt.hist(st_devs)
plt.axvline(standard_deviation(houses['SalePrice']))
```

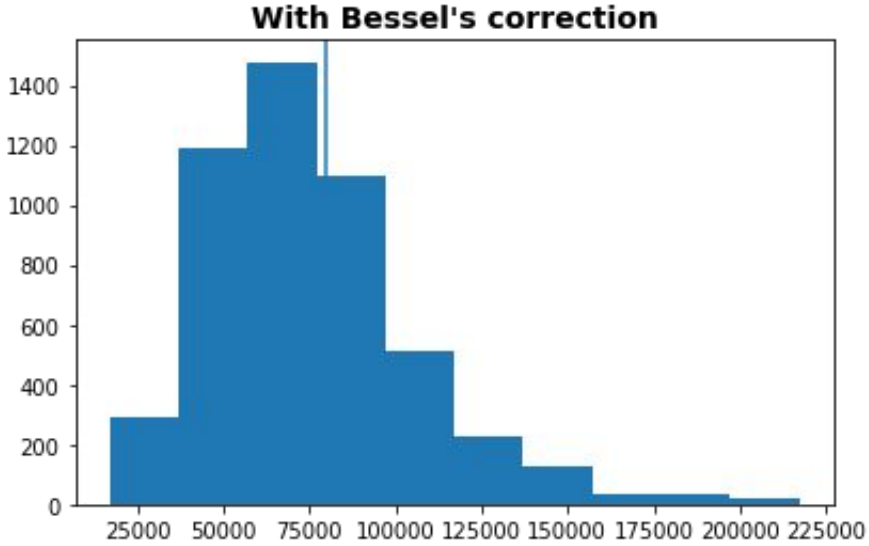
Bessel's Correction



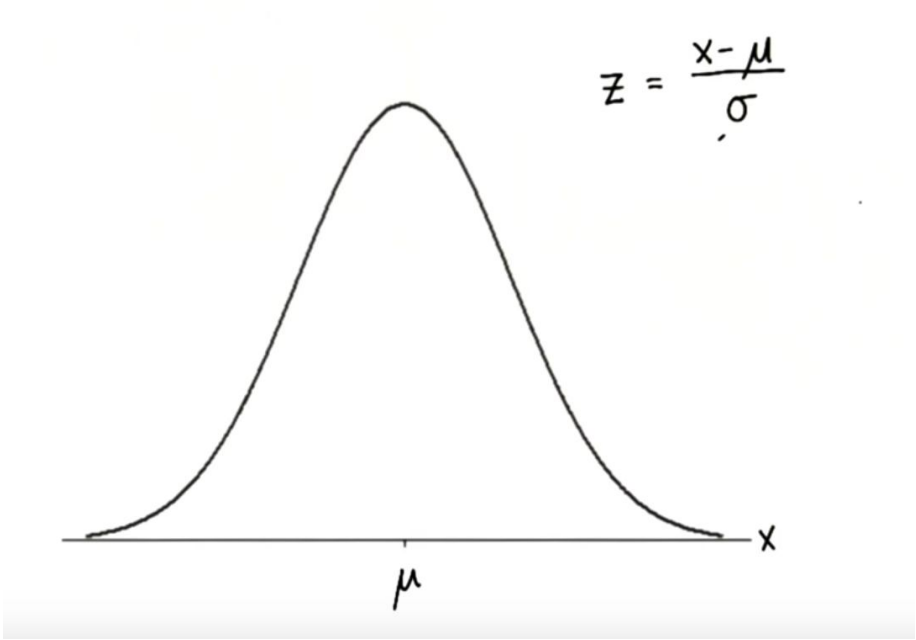
$$SD_{sample} = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}$$

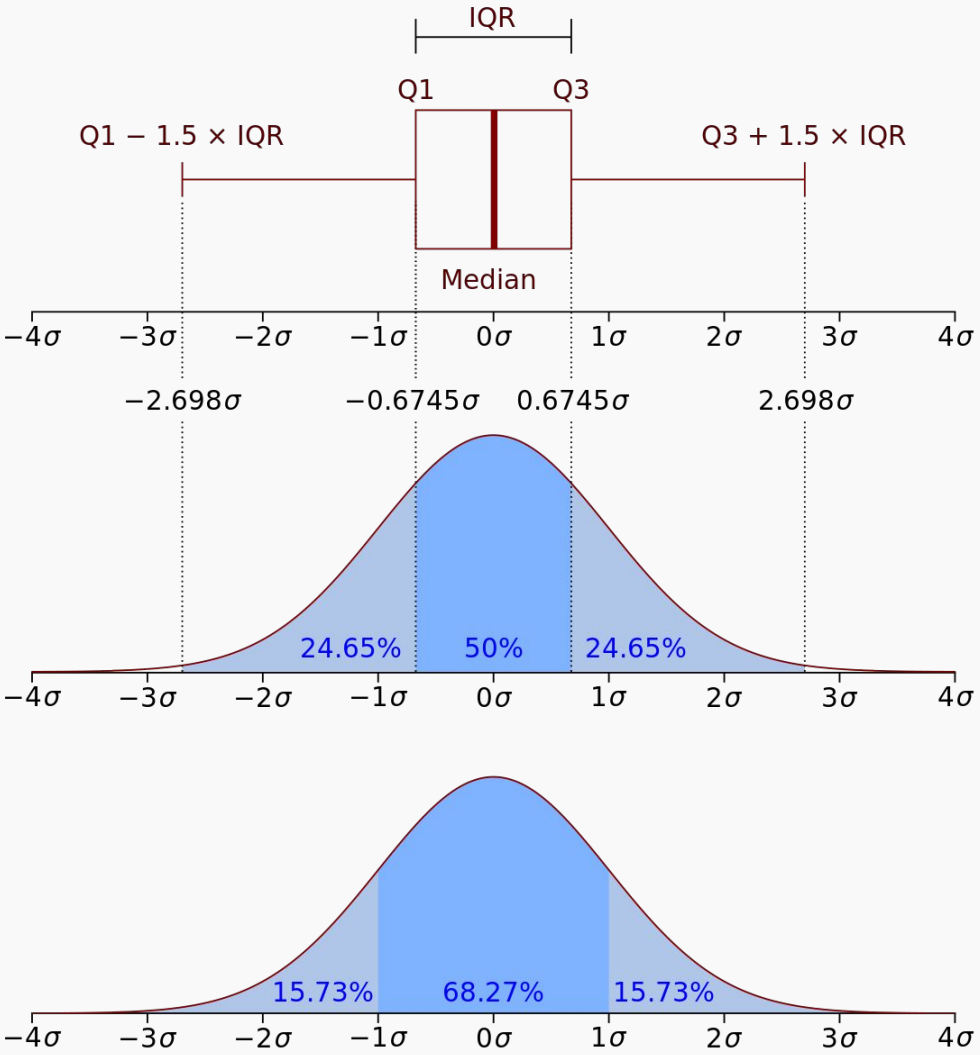
Bessel's Correction





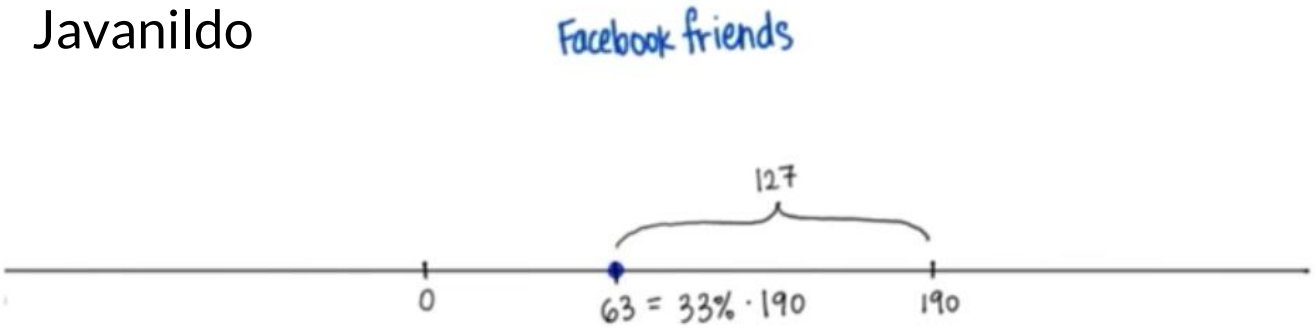
Z-Score



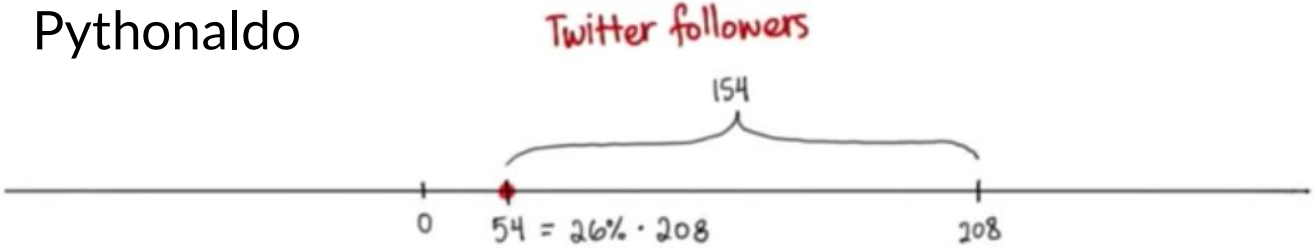

IQR
Q1
Q3
Q1 − 1.5 × IQR
Q3 + 1.5 × IQR
Median
−4σ
−3σ
−2σ
−1σ
0σ
1σ
2σ
3σ
4σ
−2.698σ
−0.6745σ
0.6745σ
2.698σ
24.65%
50%
24.65%
15.73%
68.27%
15.73%

Quiz: Quem é mais popular?

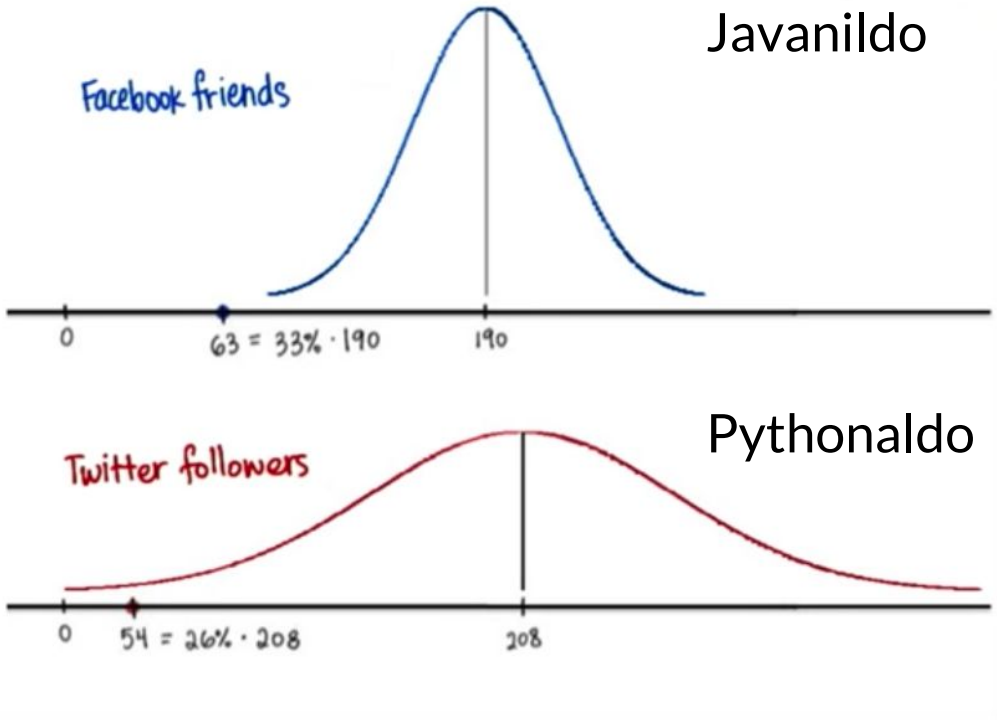
Javanildo



Pythonaldo



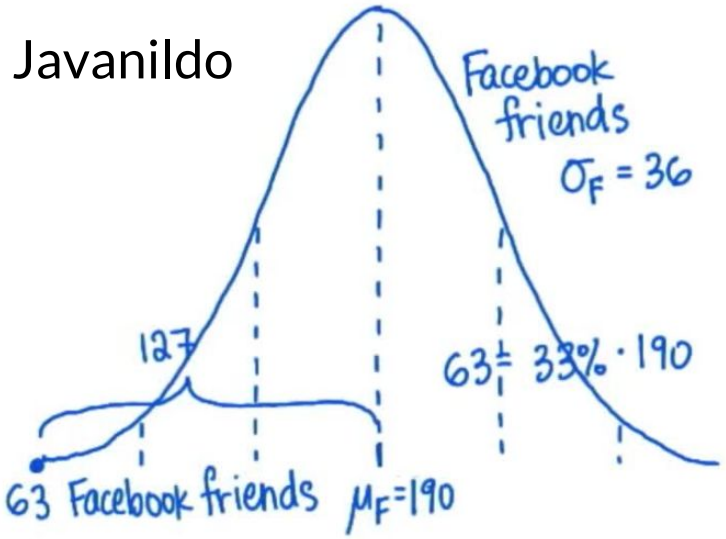
Quiz: Quem é mais popular?



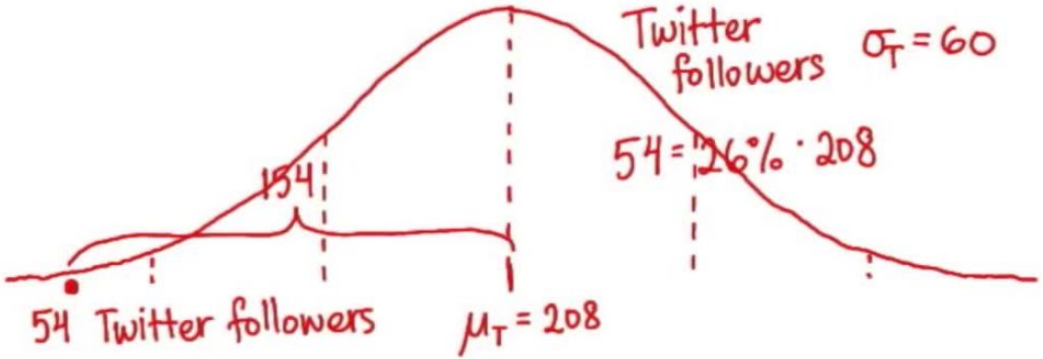
Quiz: Quem é mais popular?

Qual a distância em desvios padrões do número de amigos em relação a média?

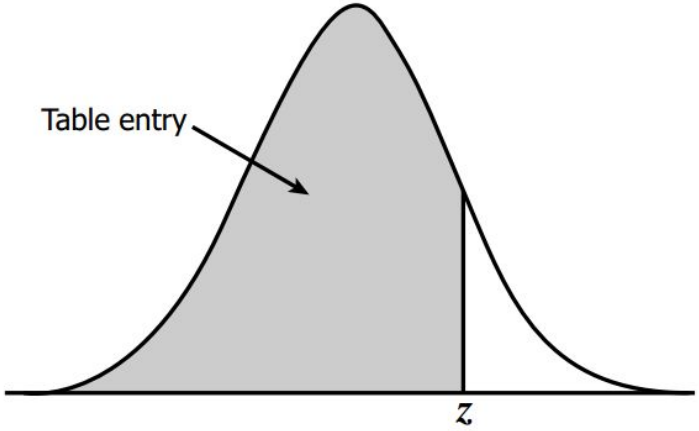
Javanildo



Pythonaldo



Z-Table



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

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

Qual a percentagem de pessoas que possuem menos de 240 amigos no facebook?


MacBook Pro