IMD0105 - Special Issues in Information Technology VI

Introduction to Statistics III

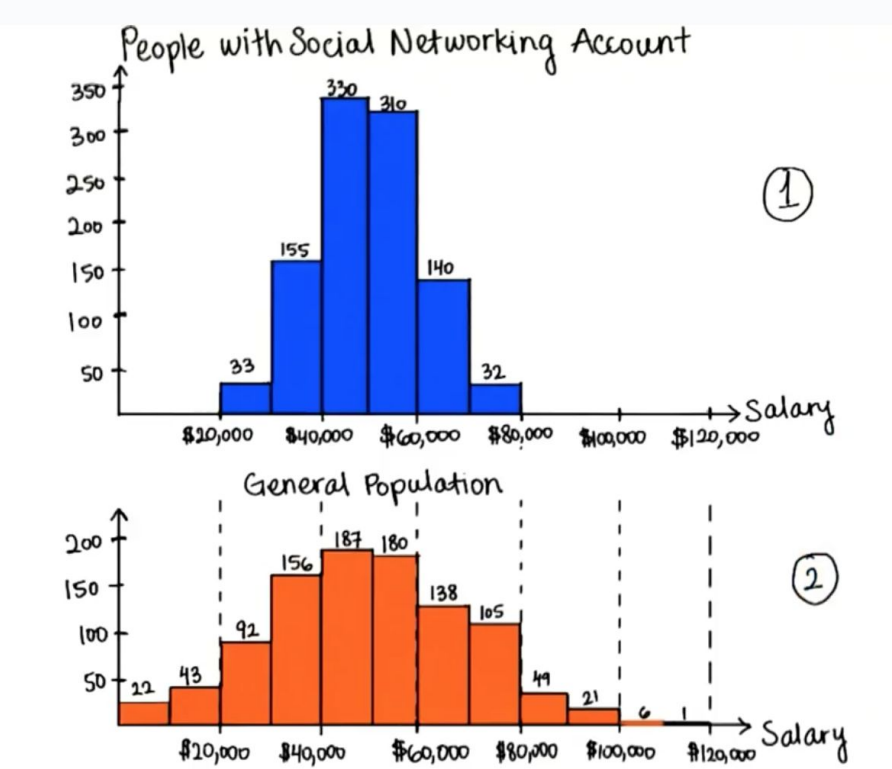
Natal-RN
April 2017

Agenda

- Quantify frequency
- Quatile
- Boxplot
- Measure variability (mean, variance, standard deviation)
- Z-Score
- Z-Table
- Facebook friends

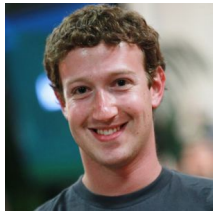
Previously on last class (...)

Social Networkers' Salaries



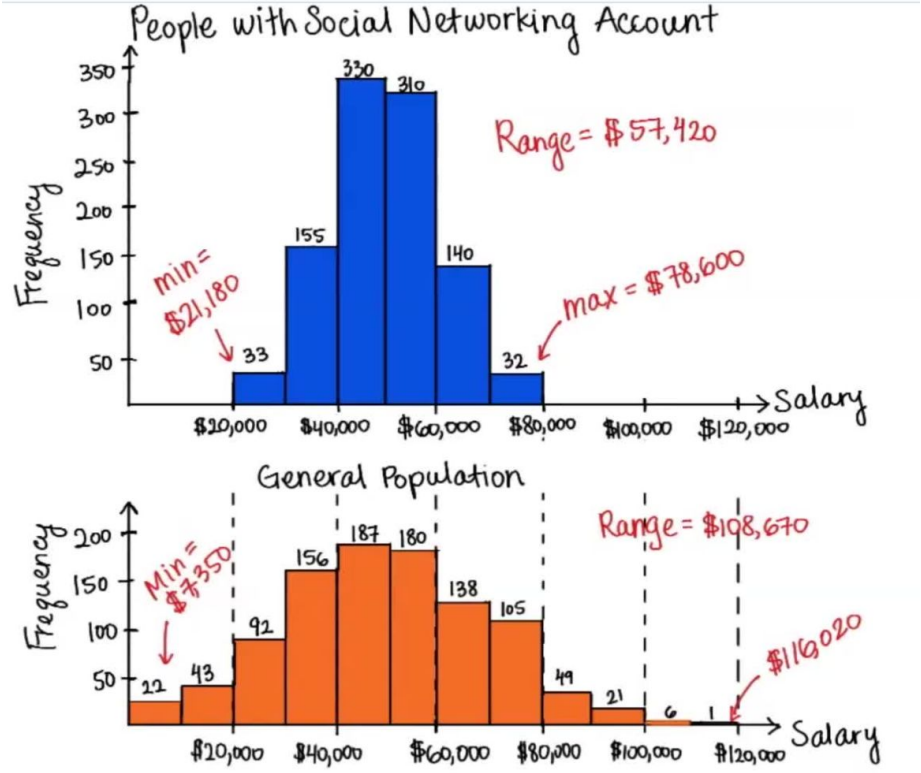
Should you get a networking account?

What's the difference between these two distributions?



$10M Salary

Quantify Spread

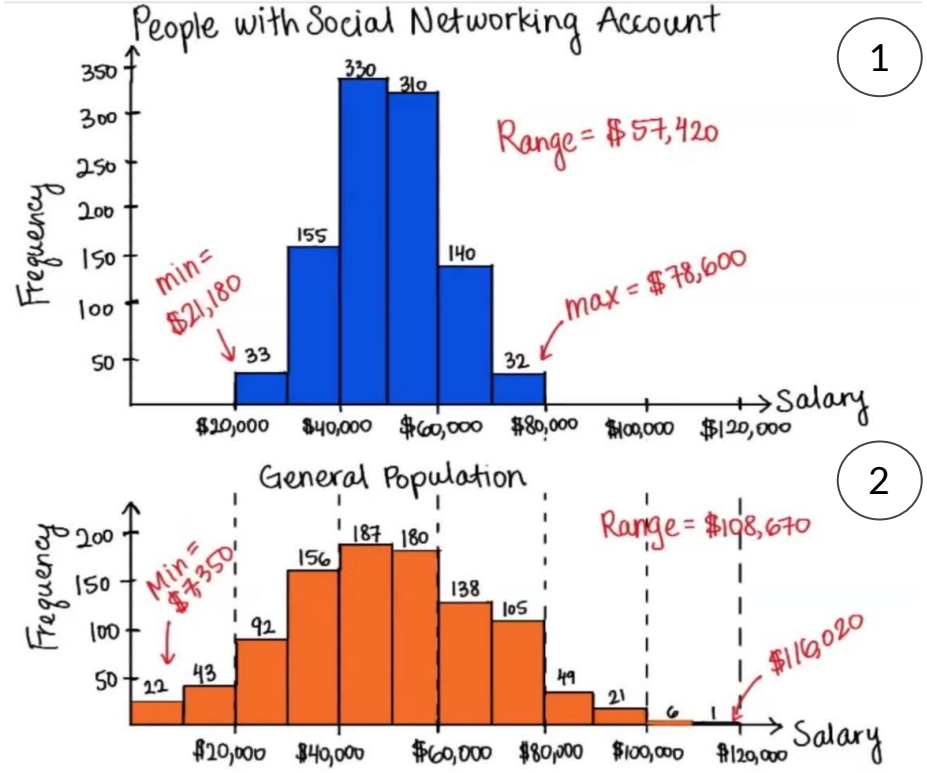


Range changes when we add new data to the dataset?

1) Always
2) Sometimes
3) Never

Quantify Spread (Quartile Q1 - Q3)

Chop of the tails



Sample 1		Sample 2	
38,946		33,219	
43,420		36,254	
49,191	Q1	38,801	Q1
50,430		46,335	
50,557	Q2 Median	46,840	Q2 Median
52,580		47,596	
53,595		55,130	
54,135	Q3	56,863	Q3
60,181		78,070	
10,000,000		88,830	

Q3 - Q1 = Interquartile range (IQR)

IQR

Sample 1

38,946
43,420
49,191 Q1
50,430
50,557 Q2 Mediana
52,580
53,595
54,135 Q3
60,181
10,000,000

Q3 - Q1 = IQR
4944

Sample 2

33,219
36,254
38,801 Q1
46,335
46,840 Q2 Mediana
47,596
55,130
56,863 Q3
78,070
88,830

Q3 - Q1 = IQR
18,062

Quiz: True or False

Room: 4PSX6ZWHG

What is an outlier?

Sample 1

38,946
43,420
49,191 Q1
50,430
50,557 Q2 Mediana
52,580
53,595
54,135 Q3
60,181
10,000,000

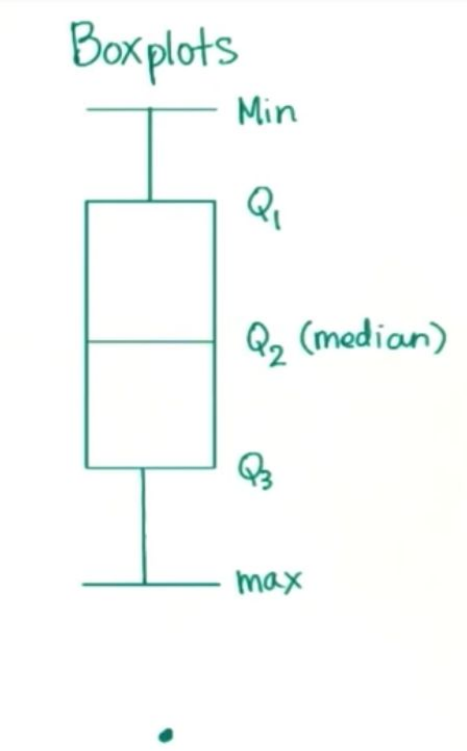
Q3 - Q1 = IQR
4944

What values do you think are outliers for this dataset?

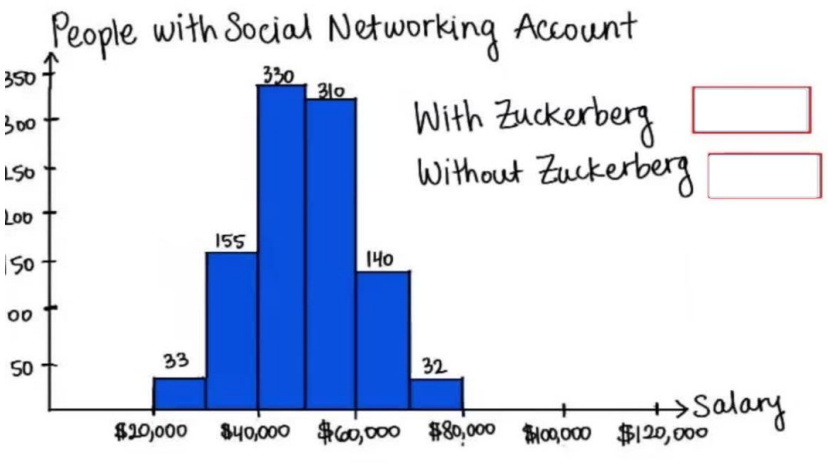
- $60,000
- $80,000
- $100,000
- $200,000

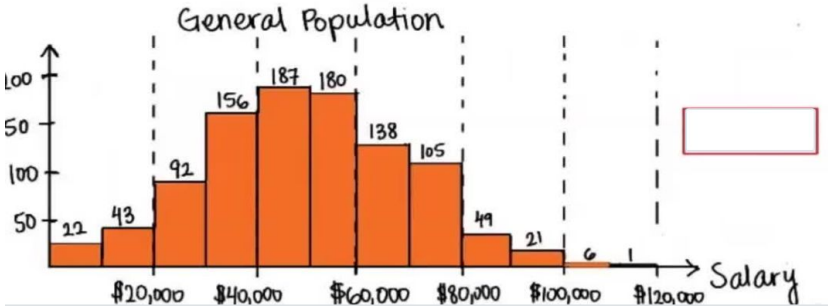
Definition:
Outlier < Q1 - 1.5 x IQR → 41,775
Outlier > Q3 + 1.5 x IQR → 61,551

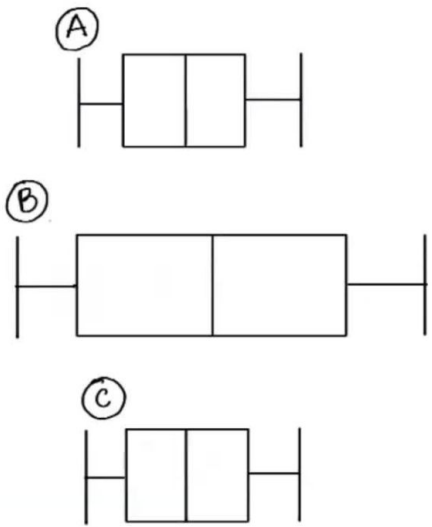
Match Boxplots



Quiz: Match Boxplots



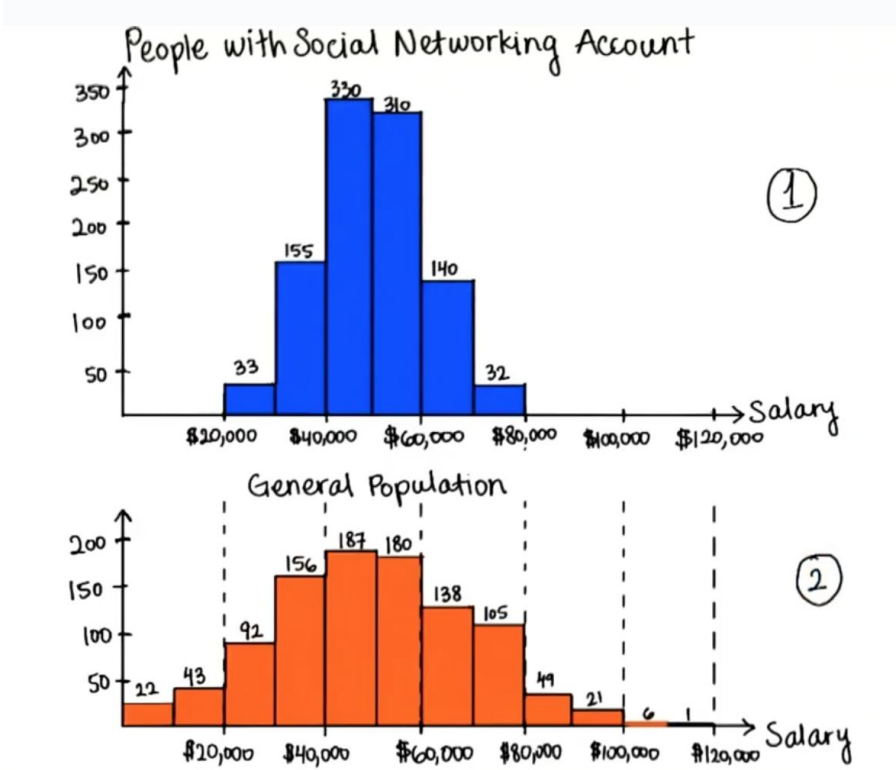


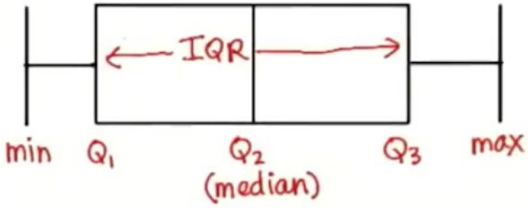


socrative
by MasteryConnect

Room:
4PSX6ZWHG

Problem with IQR





Will the mean always be between Q1 and Q3?

- Yes
- No

Coding

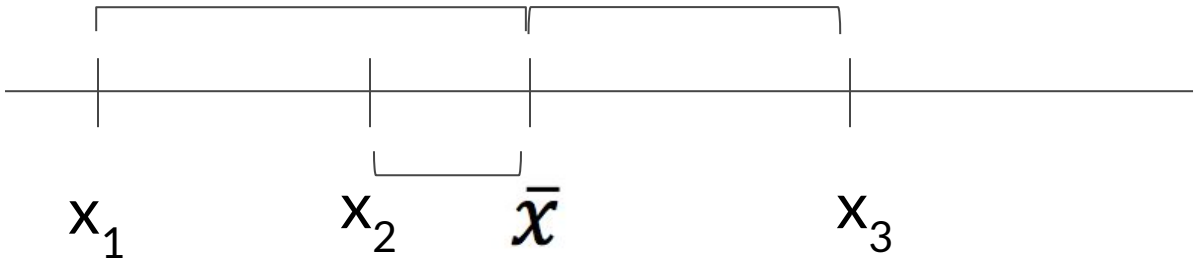
https://goo.gl/ib1F2s

Measure Variability

We need one number that describe the spread of sample that takes all the data into account.

~~Range~~
~~IQR~~

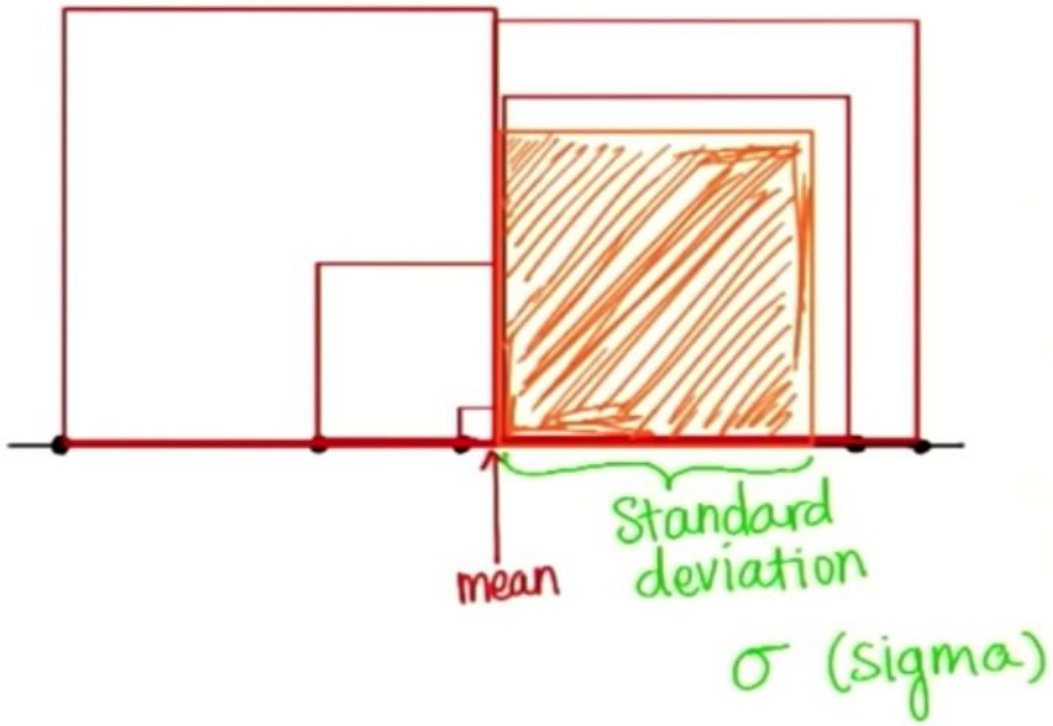
Measure Variability (idea)

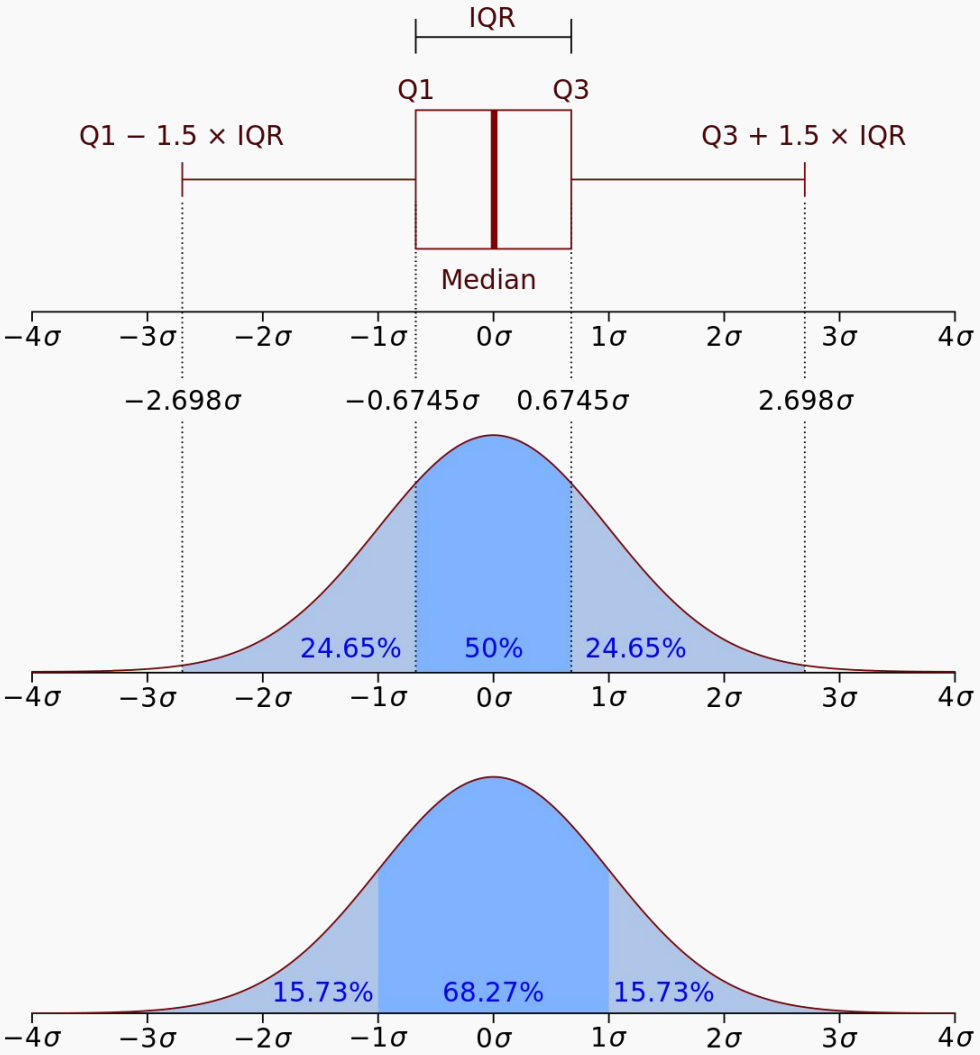


Variance $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Standard Deviation σ $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

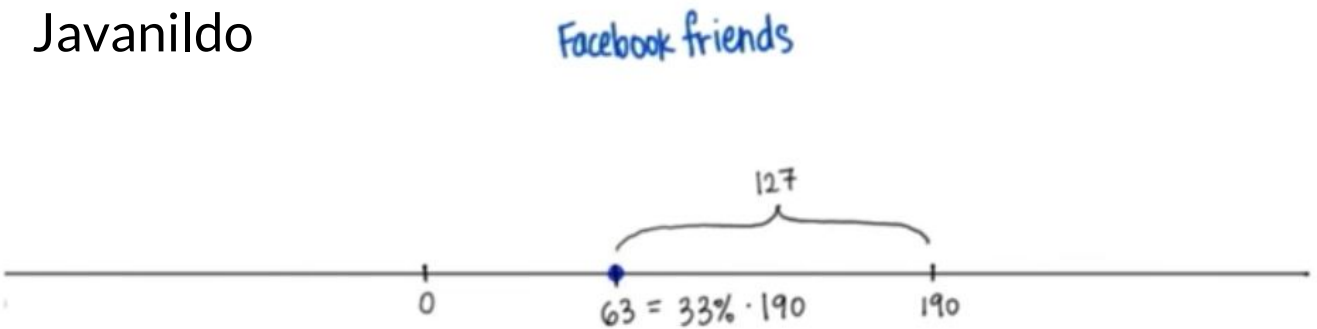
Measure Variability (idea)



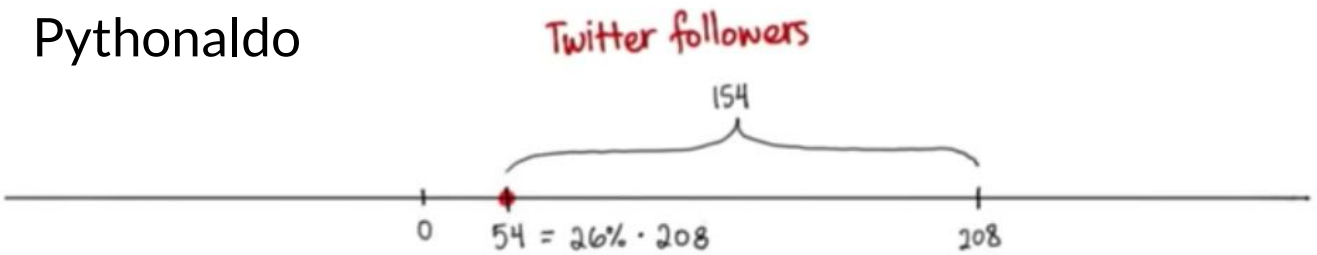

IQR
Q1
Q3
Q1 − 1.5 × IQR
Q3 + 1.5 × IQR
Median
−4σ
−3σ
−2σ
−1σ
0σ
1σ
2σ
3σ
4σ
−2.698σ
−0.6745σ
0.6745σ
2.698σ
24.65%
50%
24.65%
15.73%
68.27%
15.73%

Quiz: Who is the most popular?

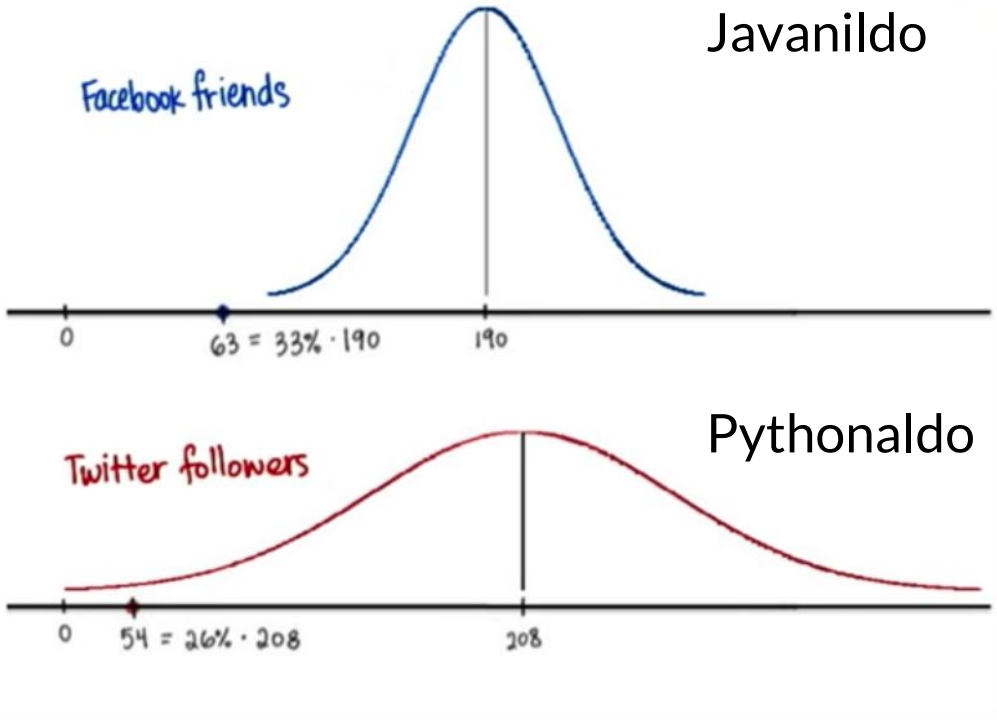
Javanildo



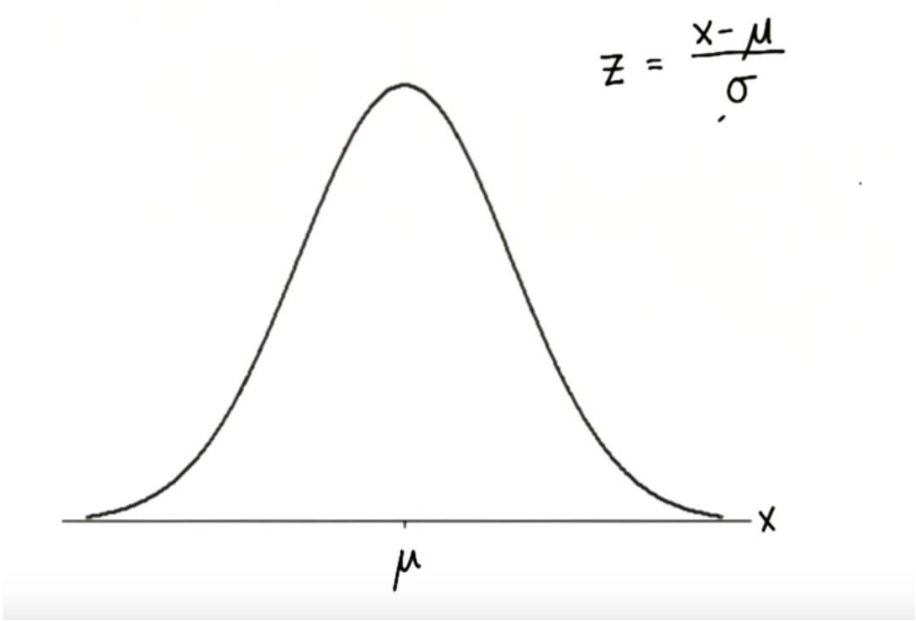
Pythonaldo



Quiz: Who is the most popular?



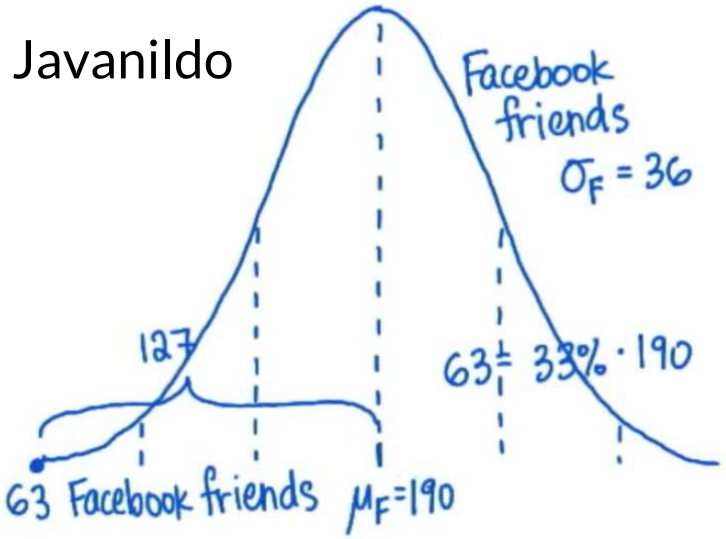
Z-Score



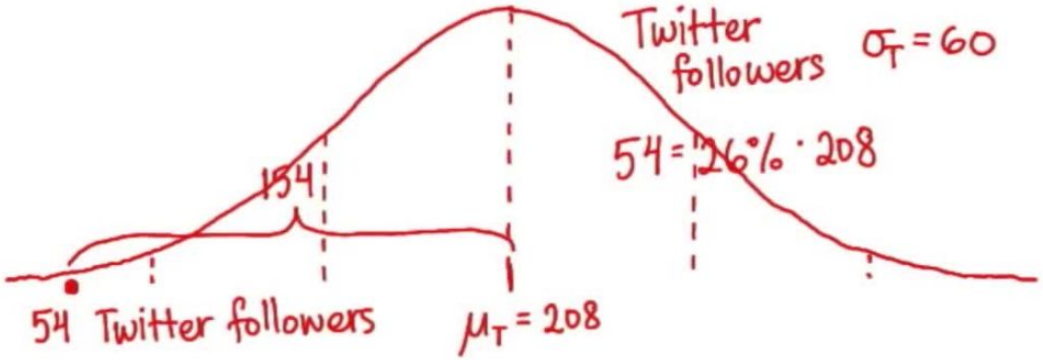
Quiz: Who is the most popular?

How many standard deviation is the number of friends from the mean?

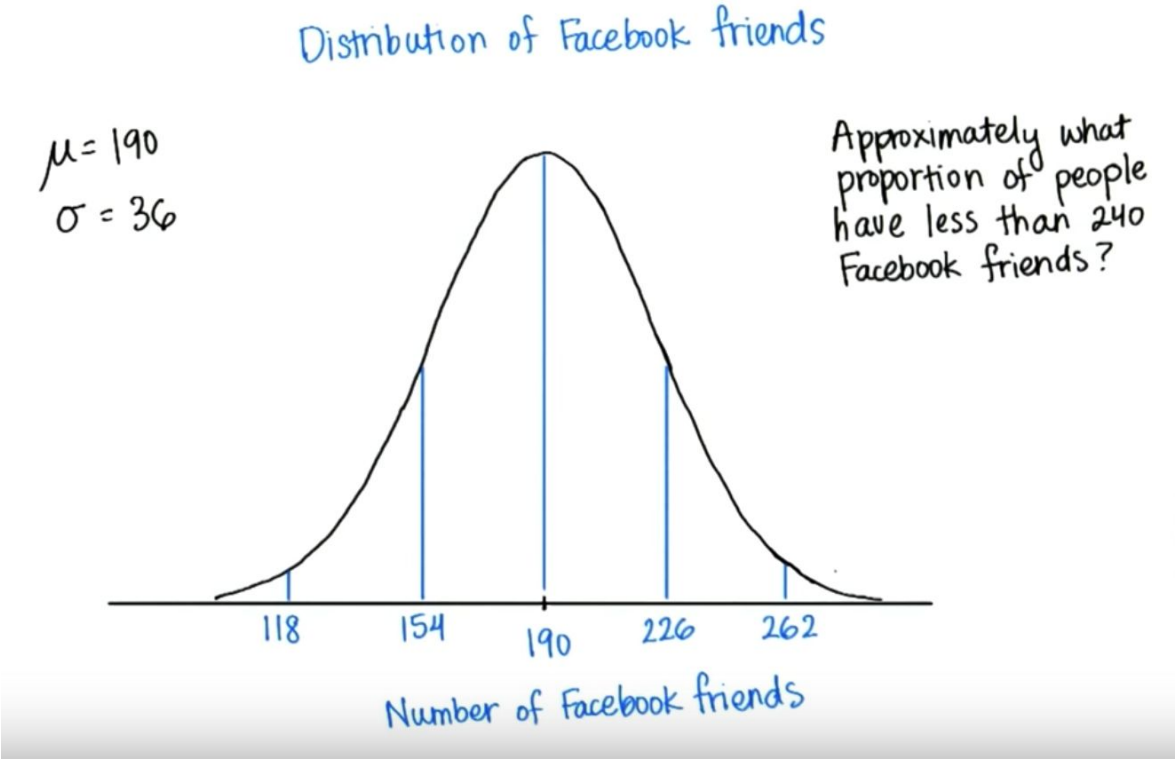
Javanildo



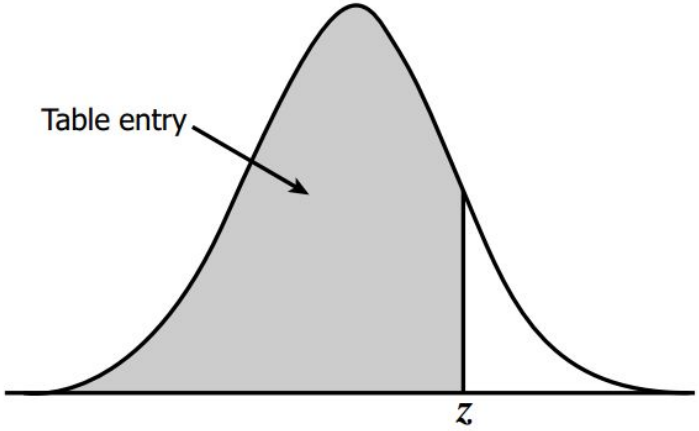
Pythonaldo



Quiz: Facebook friends (Table Z)



Z-Table



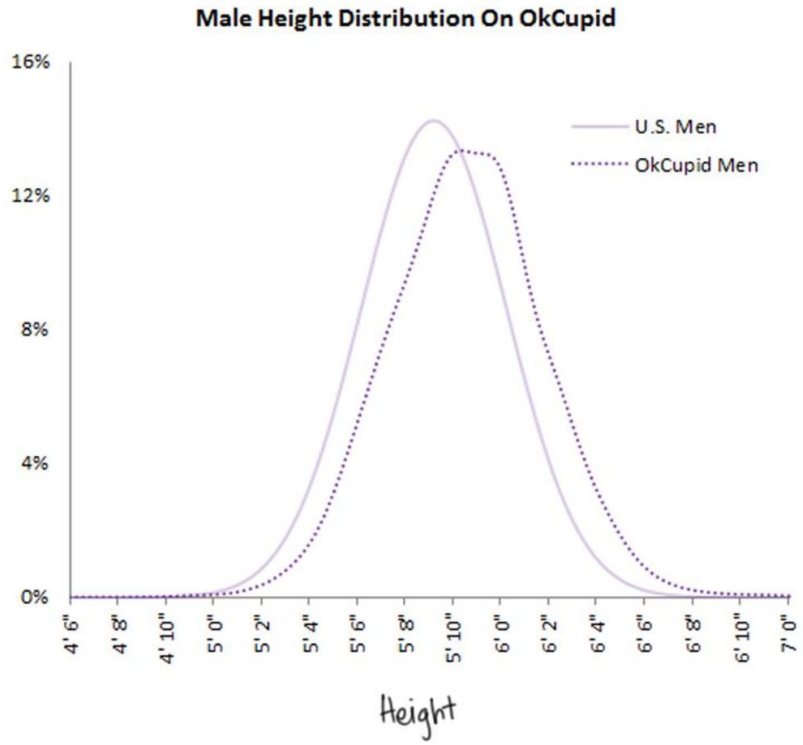
http://www.z-table.com/

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

Coding

https://goo.gl/9opGD6

Curious Quiz: OkCupid



The OkCupid blog shared this graph. What is going on here?