Essential Statistics 101

sketchwow

Outline

- Statistics?
- Sampling
- Descriptive Stats
  1. central tendency
  2. spread
  3. position
- Point vs. Interval
- Distribution (Normal)

What is Statistics?

population
70 million people
58%

Simple Random Sampling

60% — sample n=10000
Margin Error
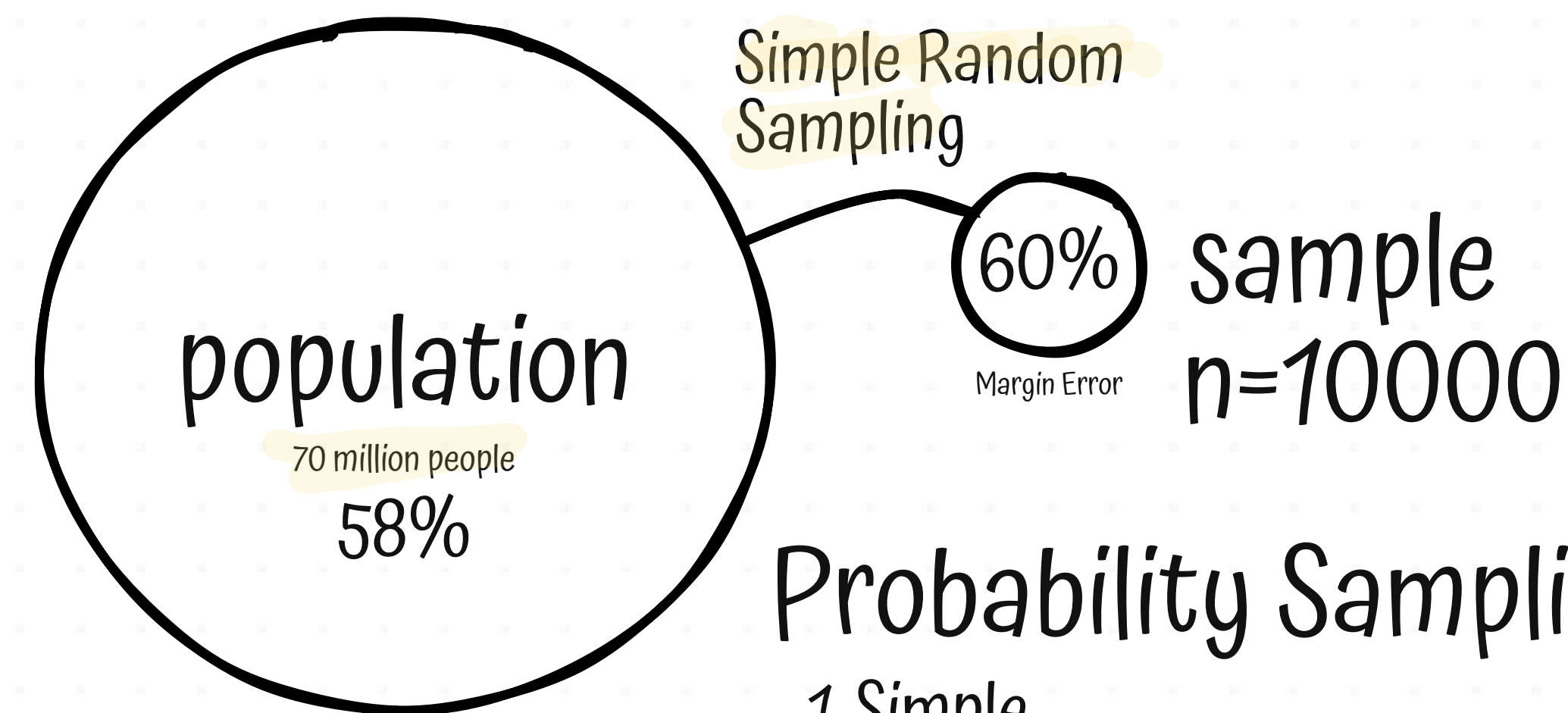
Constraints
1. Budget
2. Timeline

CPI = Cost Per Interview
200K, CPI=500, N= 400
MOE = 5-6%

Probability Sampling
1. Simple
2. Systematic
3. Stratified

Bootcamp — sampling → sample — generalization

is this a good bootcamp?
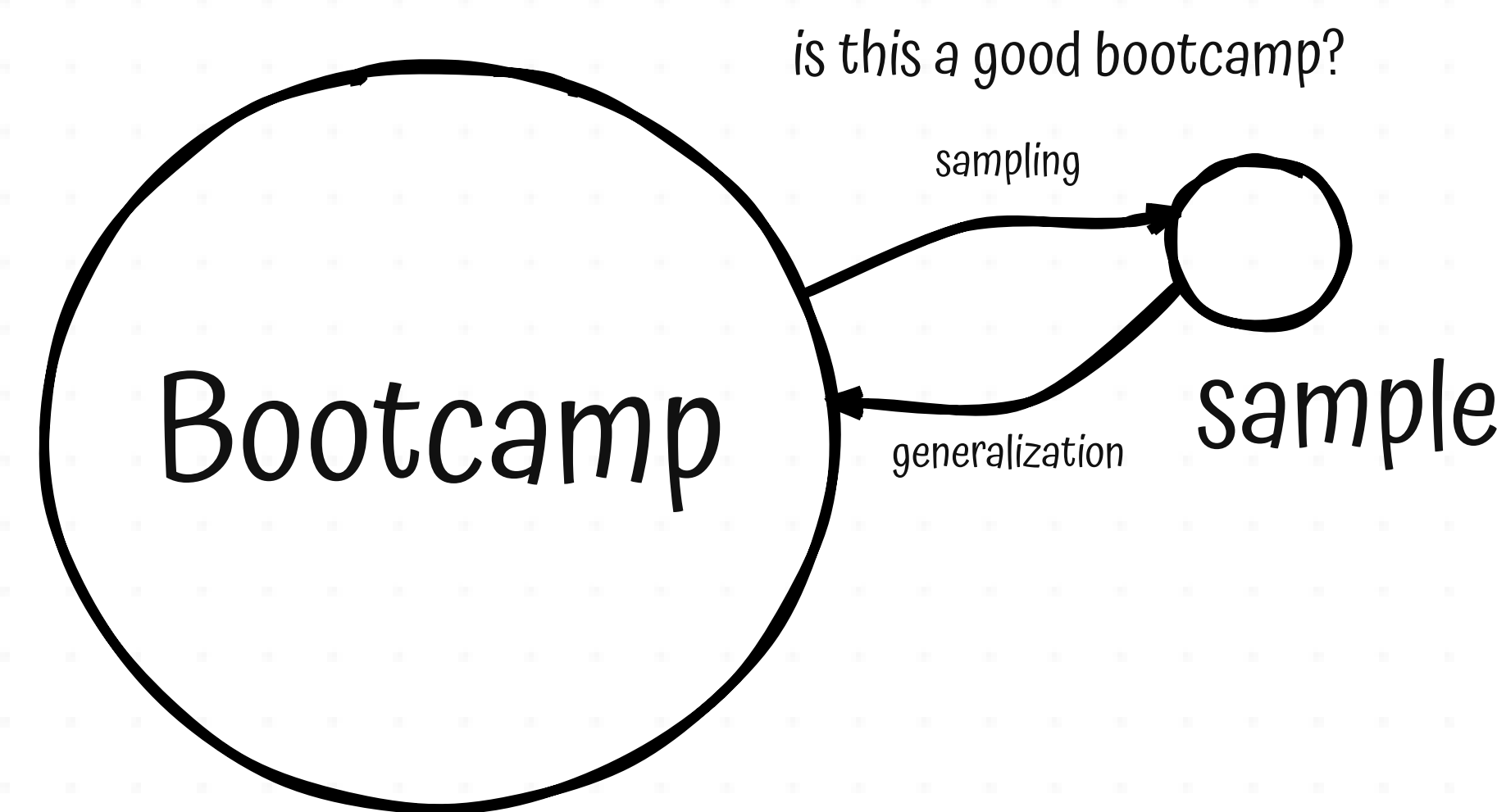
Central Limit Theorem

Population 3932
μ มิว mu

3%
6%
10% S1
9% S2
12%
12%
9%

n=400
Point Estimate

CTL holds true when
1. sample size >= 30
2. Random sampling

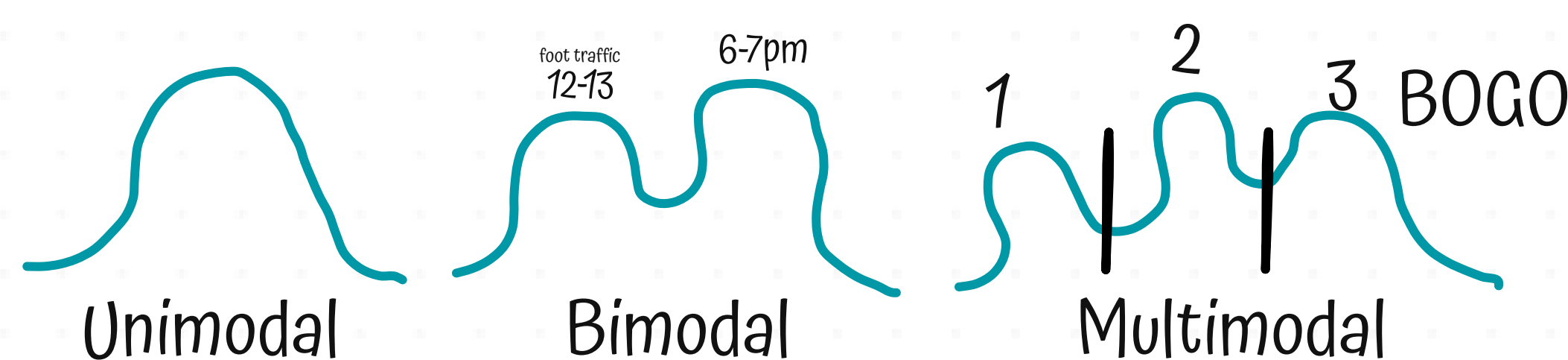## Descriptive Statistics

3 broad categories
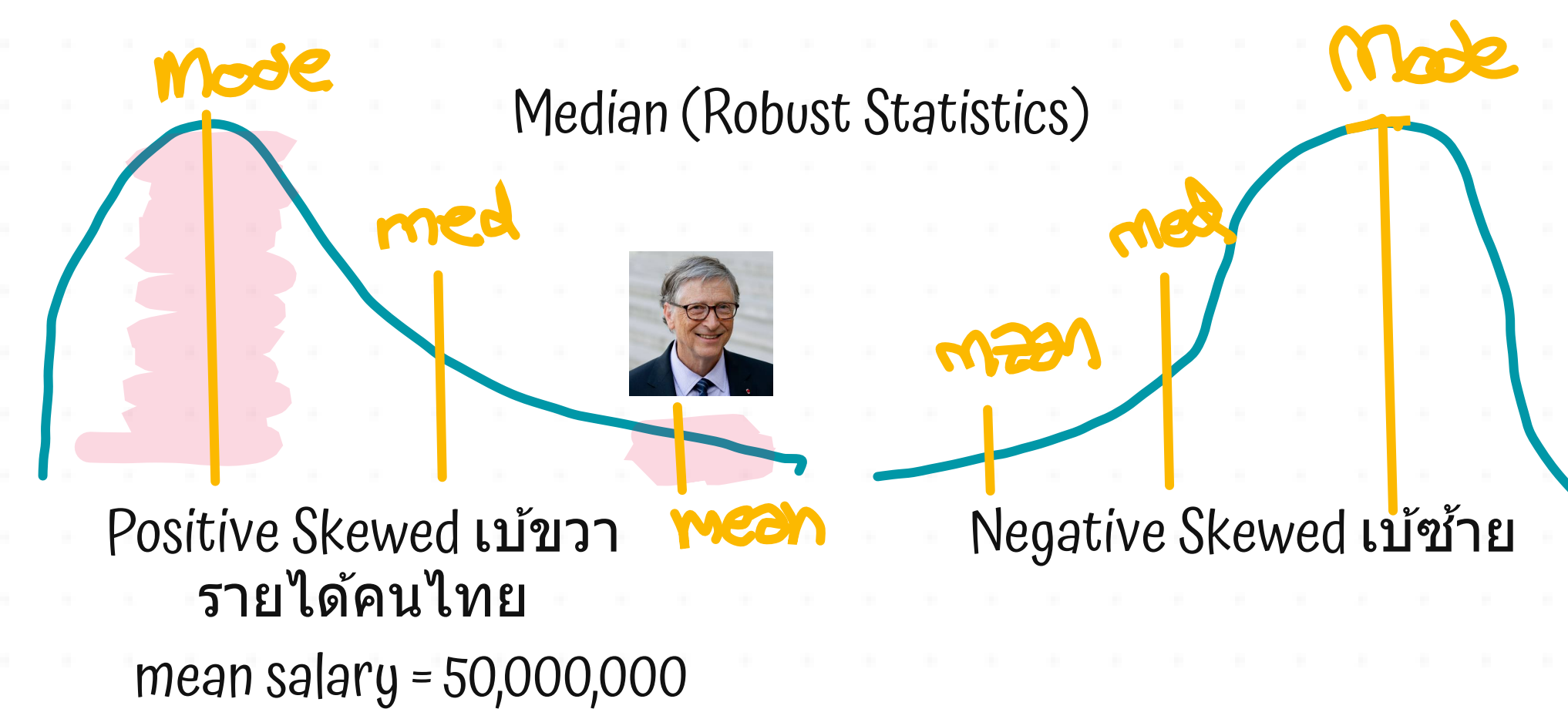
1. Measures of Central Tendency
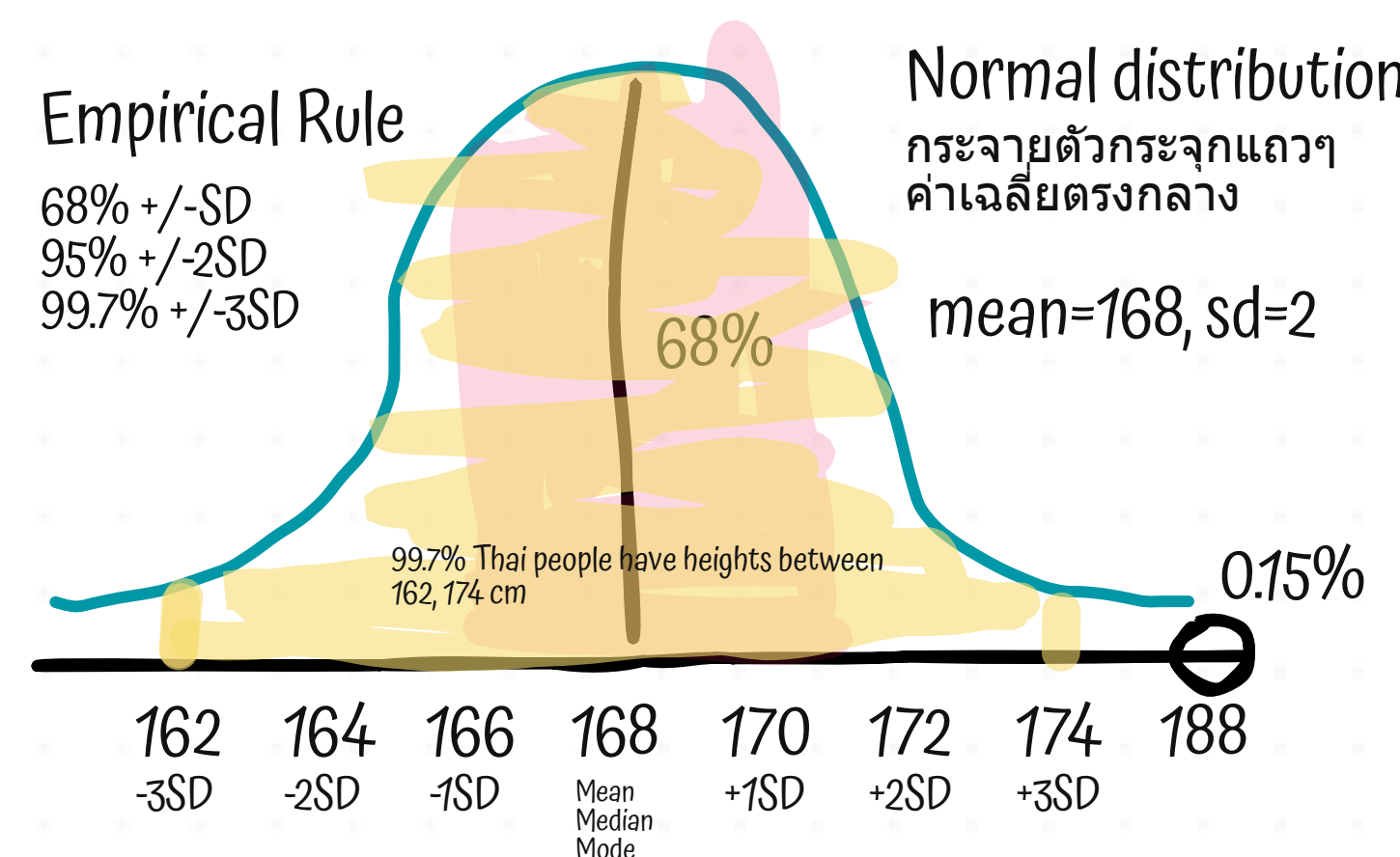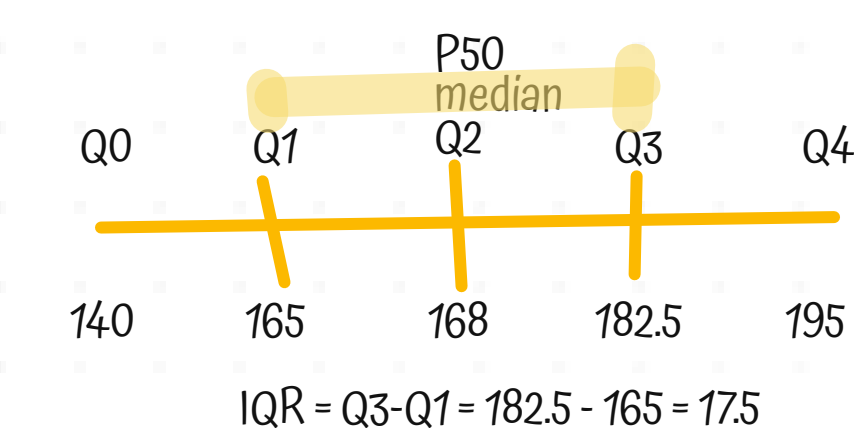   - mean
   - median
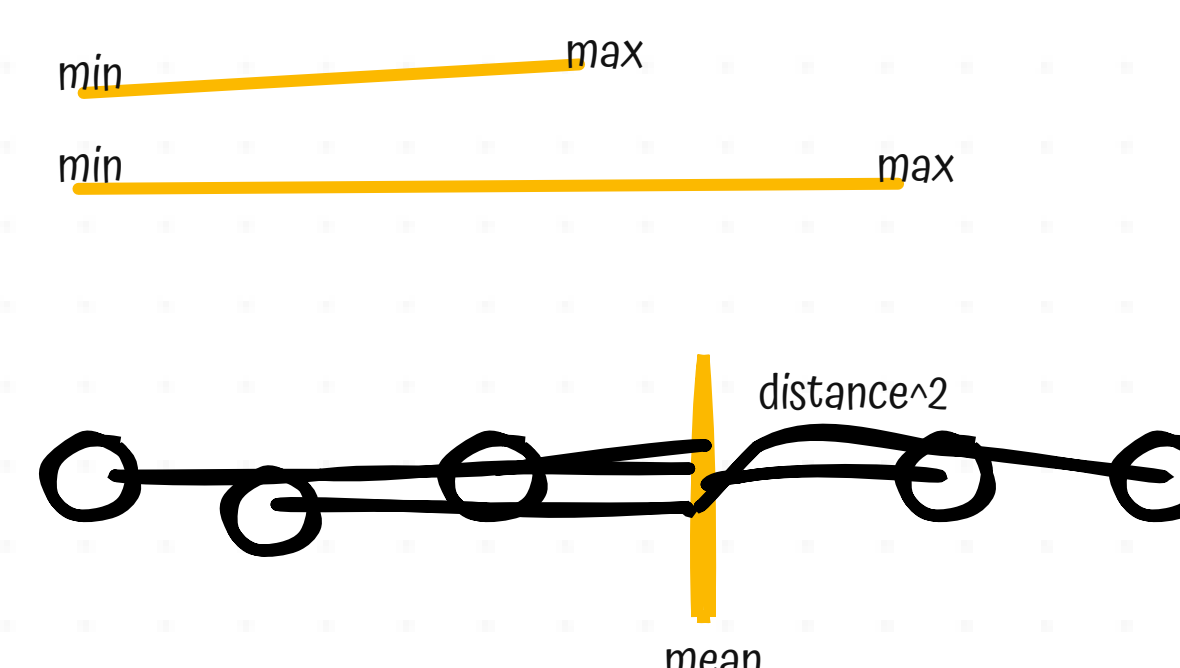   - mode
2. Measures of spread (variability)
   - standard deviation (sd)
   - variance
   - range
   - IQR
3. Measures of Position
   - Min, Max
   - Percentile (50th, 78th)
   - Quartile (Q1, Q2, Q3, Q4)

Empirical Rule
68% +/-SD
95% +/-2SD
99.7% +/-3SD

68%

Normal distribution
กระจายตัวกระจุกแถวๆ
ค่าเฉลี่ยตรงกลาง

mean=168, sd=2

99.7% Thai people have heights between 162, 174 cm

0.15%

162	164	166	168	170	172	174	188
-3SD	-2SD	-1SD	Mean Median Mode	+1SD	+2SD	+3SD	

P50 median

Q0	Q1	Q2	Q3	Q4
140	165	168	182.5	195

IQR = Q3-Q1 = 182.5 - 165 = 17.5

How to detect outliers?
1. Boxplot
2. Formula: Q3+3*IQR, Q1-3*IQR

Unimodal

Bimodal
foot traffic
12-13   6-7pm

Multimodal
1 2 3 BOGO

Median (Robust Statistics)

Mode
med
mean

Positive Skewed เบ้ขวา
รายได้คนไทย
mean salary = 50,000,000

mean
med
Mode

Negative Skewed เบ้ซ้าย

min —— max
min ———— max

distance^2
mean

Sampling Distribution of mean

μ