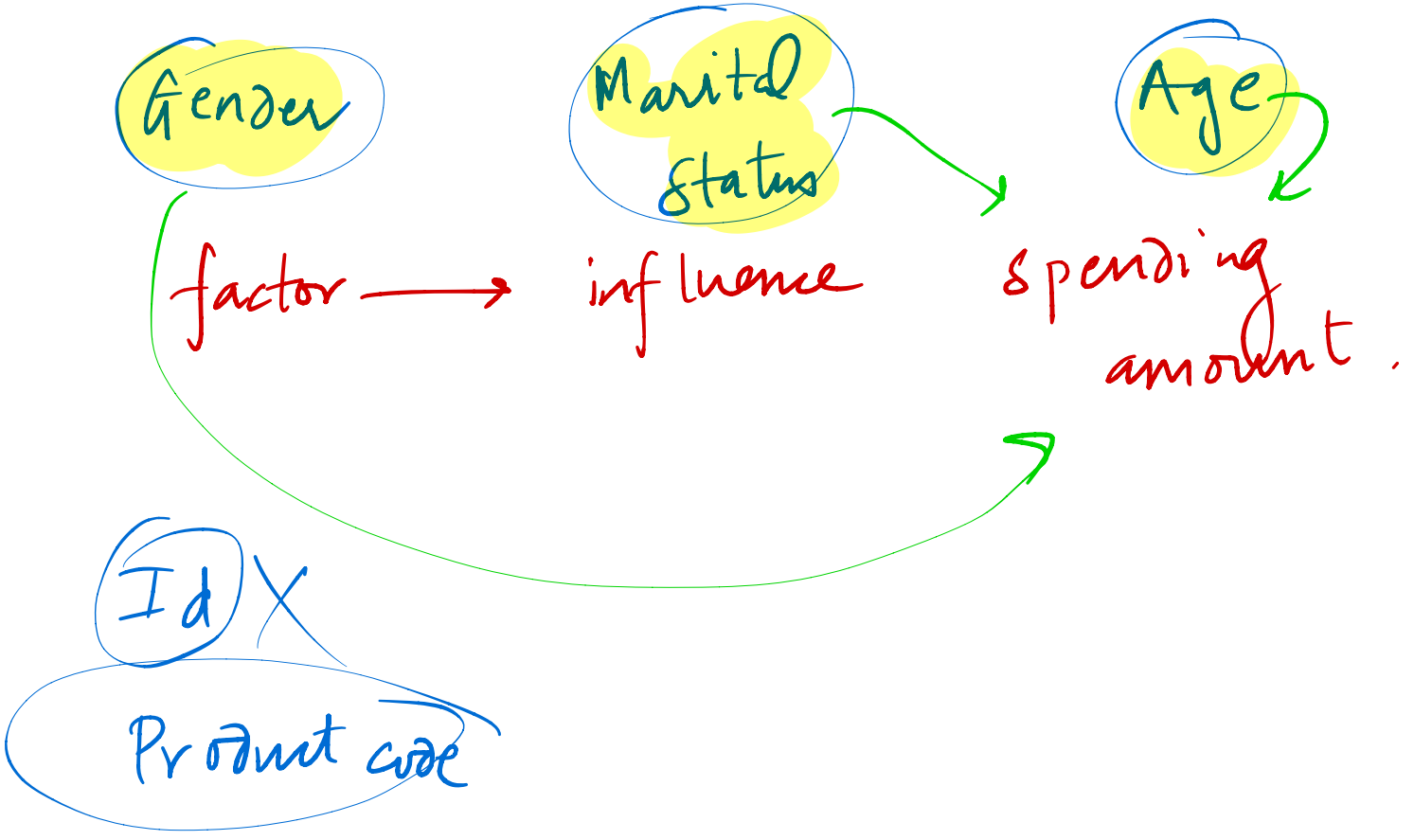
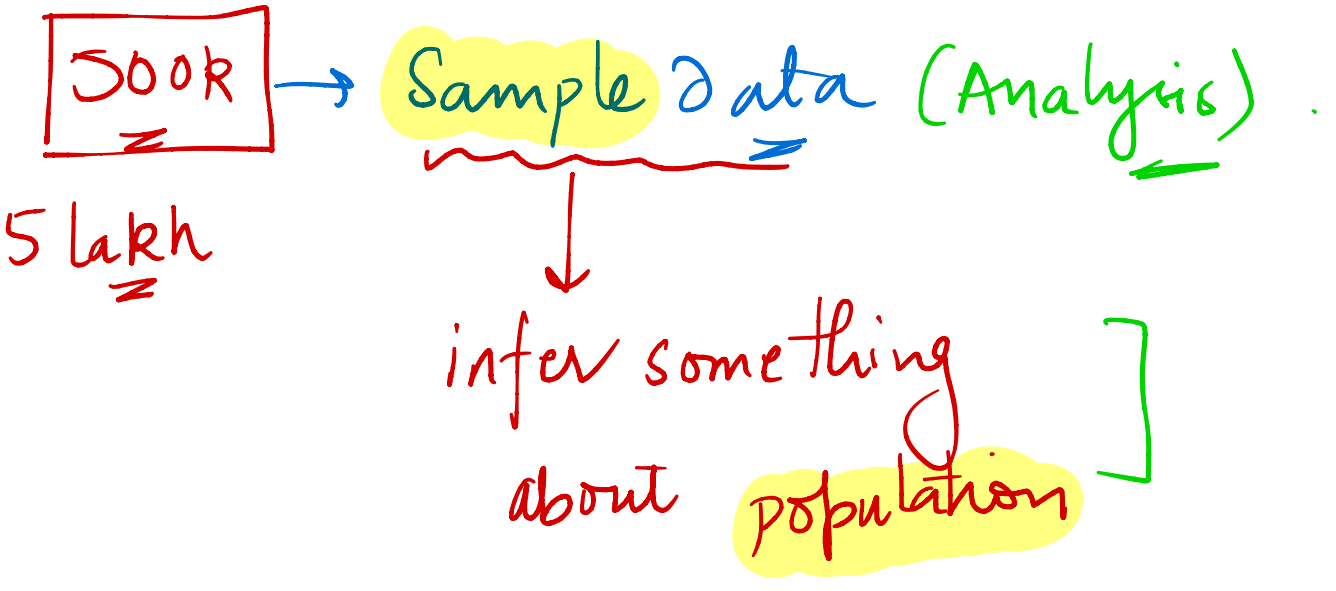
20th jan 2023

Wallmart Business Case study – Review

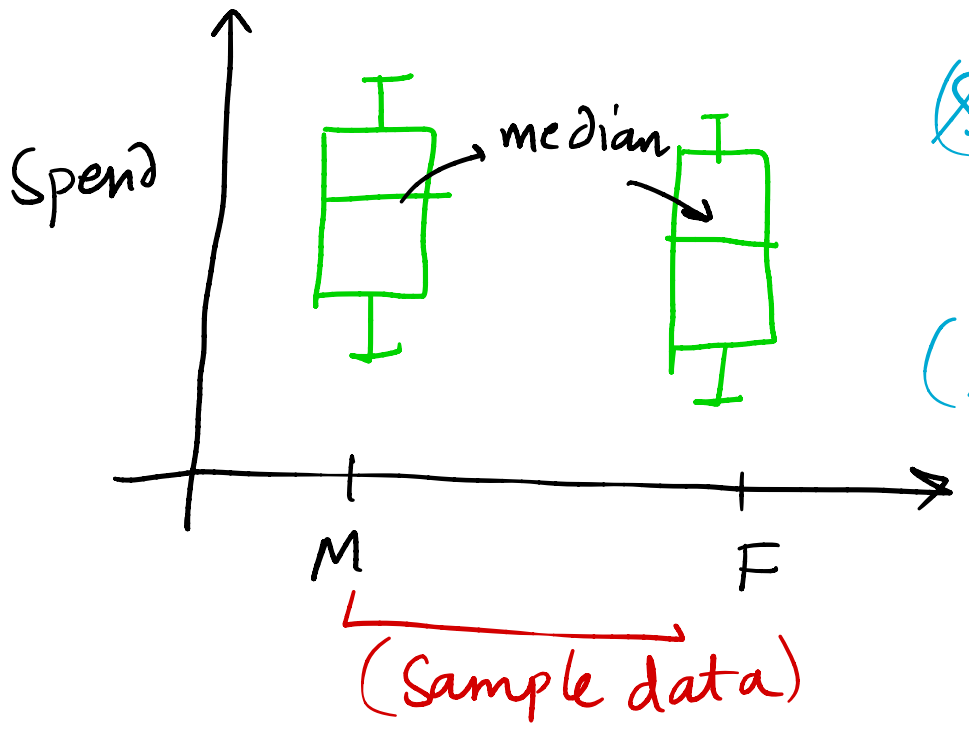
Welcome Everyone

let's start @ 9:05


Gender
Marital status
Age
factor
influence
spending amount.
Id
X
Product code


500K
Sample data (Analysis)
5 lakh
infer something
about population
10K *
[8K, 12K]
95% CI

* Gender ✓ (Analyze spend)



(Spending) male
>
(Spending) female

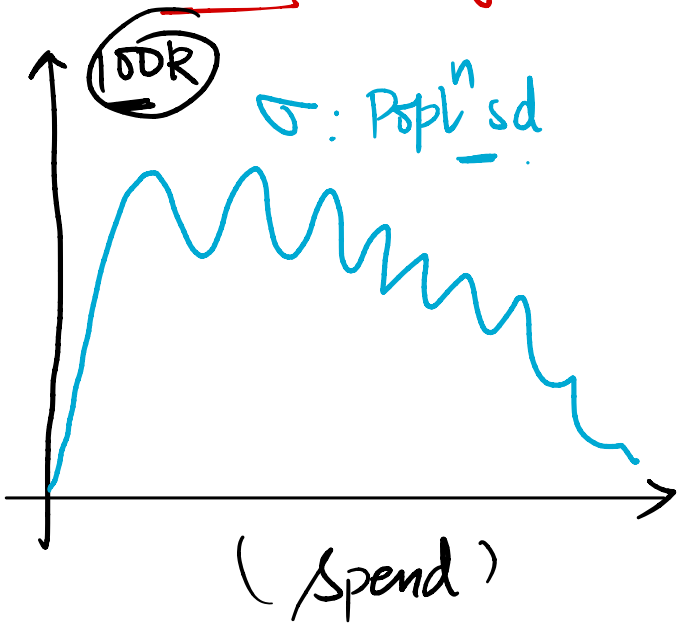
We can't assume same for population.

CLT : Central limit theorem :

* estimate of population mean from sample (μ)

mean ($\bar{x}$)

⇒ Bootstraping



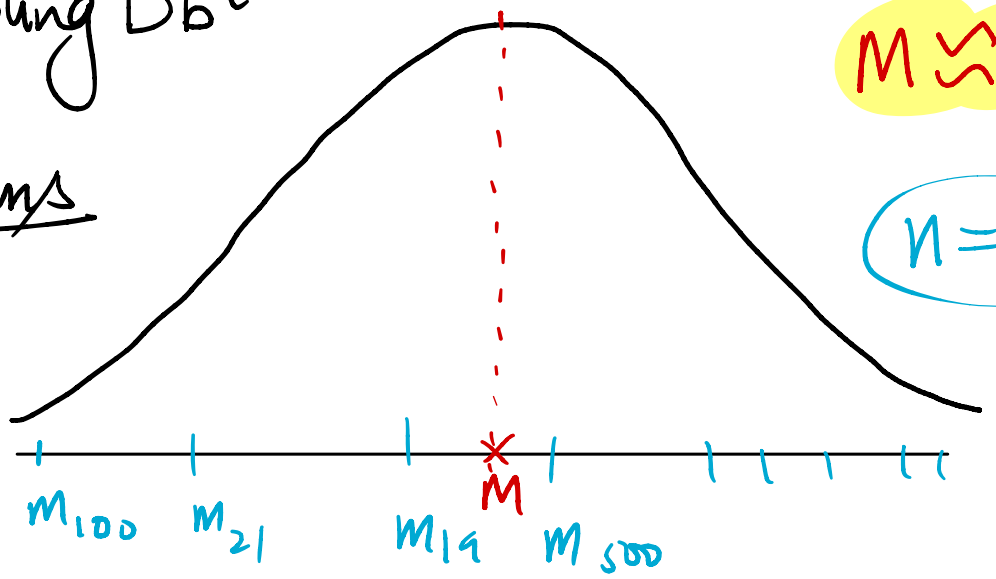
Sample size

S_1 $n=300$ → m_1 ✓

S_2 $n=300$ → m_2 ✓

S_{1000} $n=300$ → m_{1000} ✓

Sampling Dbn of Means



$M \approx \mu$ → Population Mean

$n = 300$

Mean of Sample means → ? → $\frac{m_1 + m_2 + \cdots + m_{1000}}{1000}$ → M

Standard deviation (Standard Error)

$= \sigma/\sqrt{n}$ ×

$M_{\bar{x}}$ → Mean of Sample means

$s_{\bar{x}}$ → Standard Error → Std. dev^{n} of Sample means.

.mean → np.std( )

=

$M \approx \mu$ → Pop. mean

with 95% Confidence

90%.

99%. } Interval

=

800 → ~~90%~~ confidence

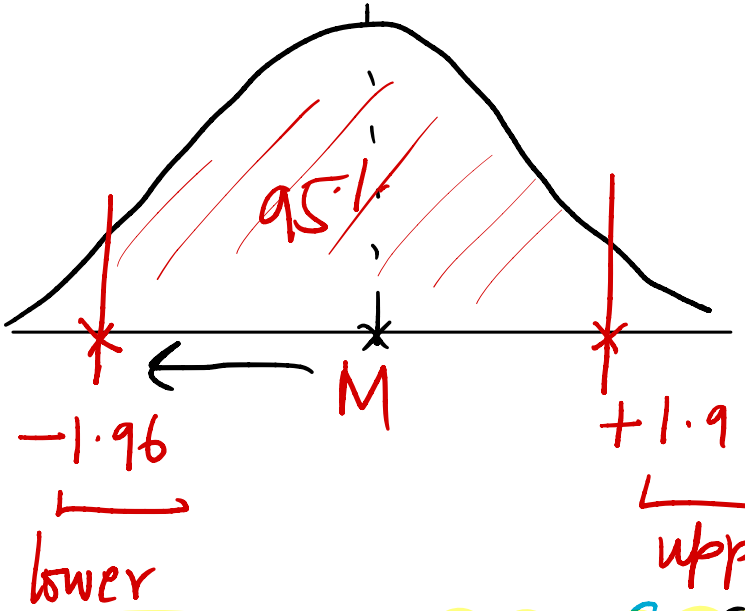
→ [700, 900] → 90% Conf

Range [700, 900] → Confidence Interval

95% C.I

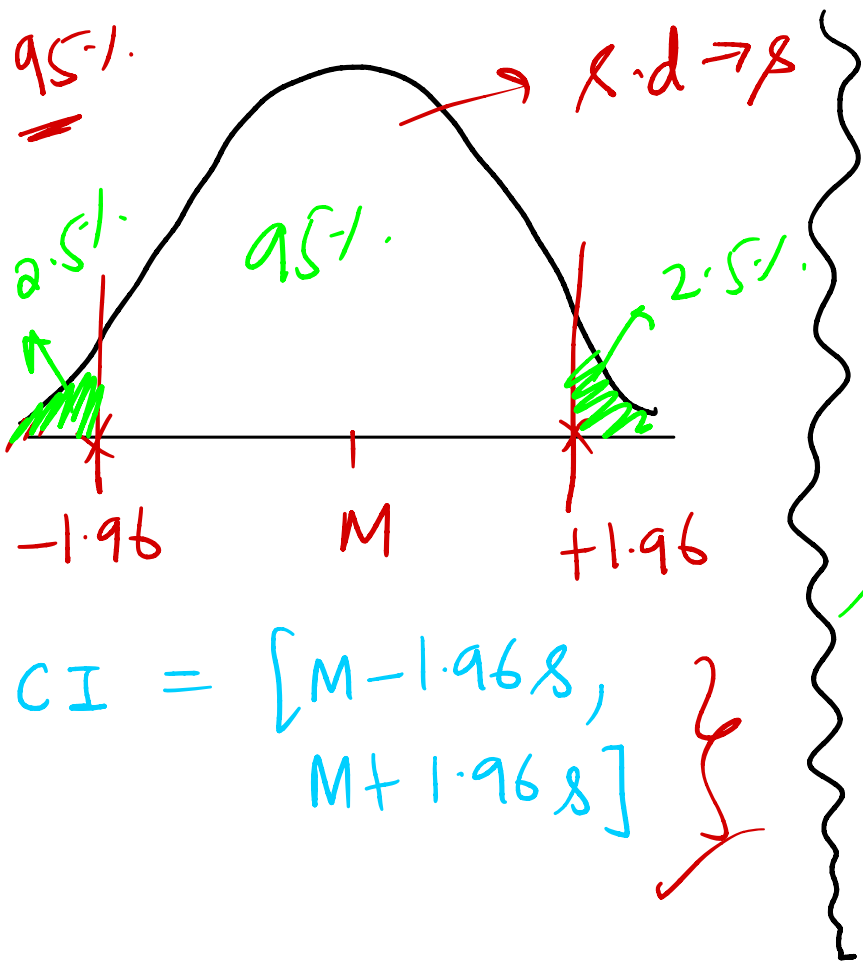
M → mean.

'S' → S.d.



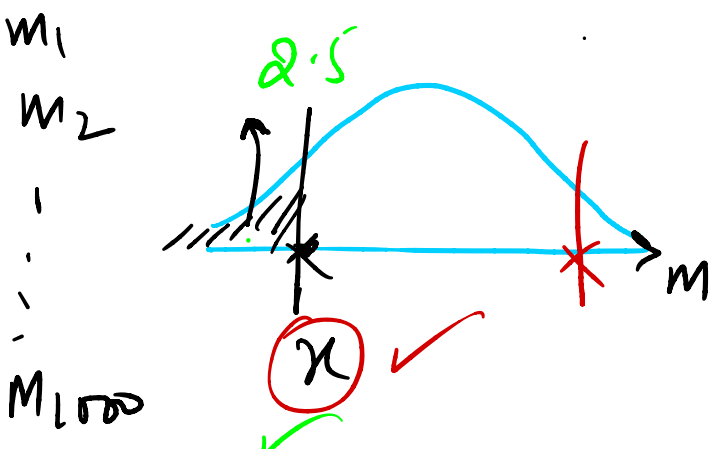
[lower, upper] → [M − 1.96 * S, M + 1.96 × S]

↳ C.I. with 95% Confidence



$\bar{x}$ · d → β

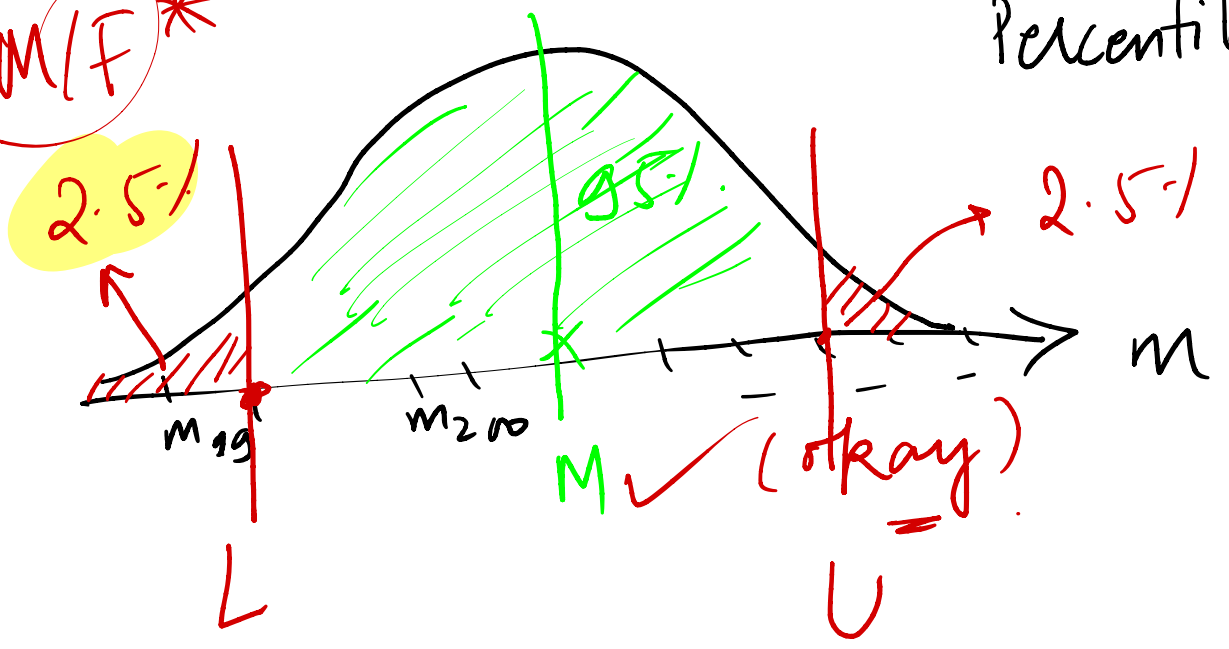
$CI = [M - 1.96\beta, \; M + 1.96\beta]$



len (df[spend < $\bar{x}$]) / total length

M/F *

Percentile method



[L, U]

- np.percentile(data, 2.5) → L
- np.percentile(data, 97.5) → U

} 95% CI

Working day (0/1)

Categorical

Count

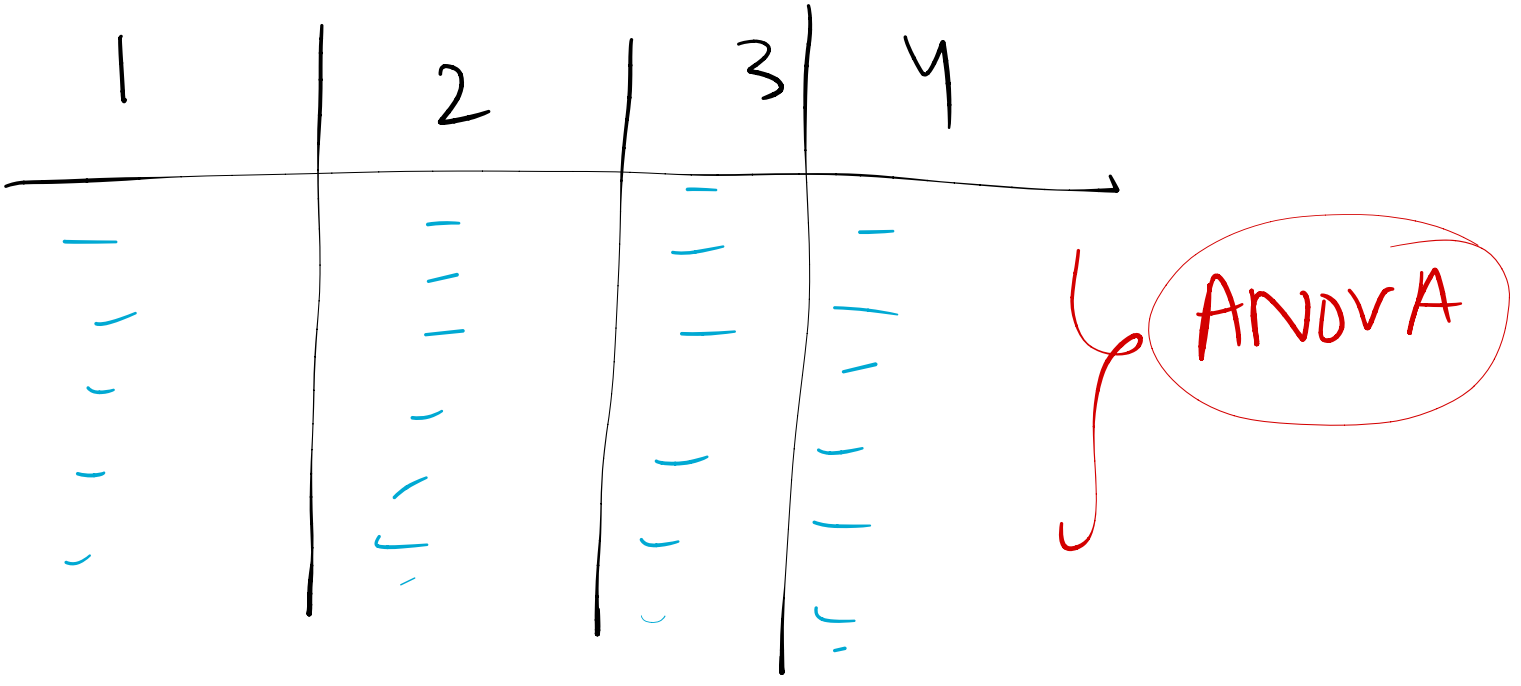
Numerical

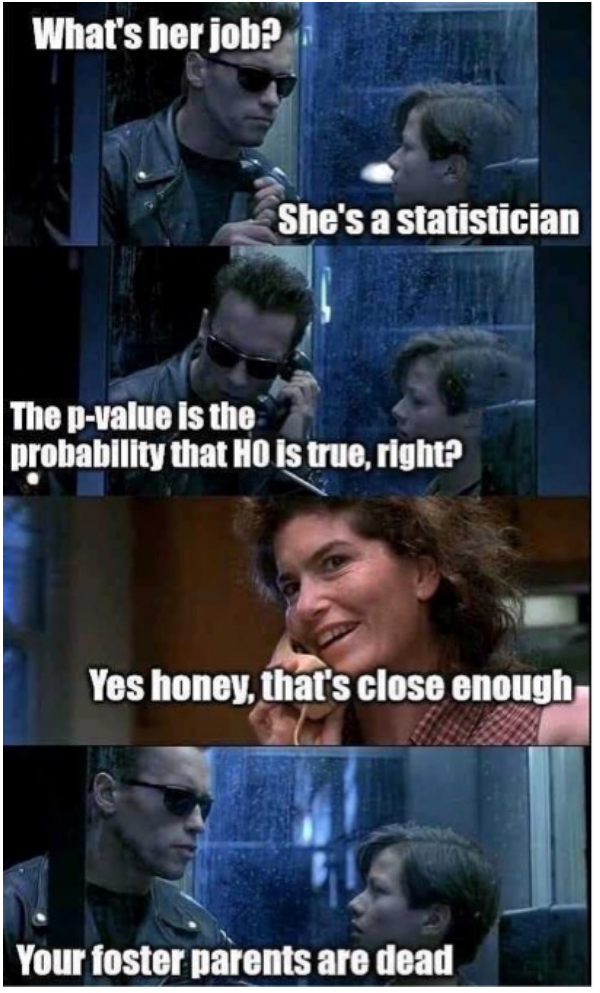
0	1
52	49
53	53
51	47
25	95
⋮	⋮

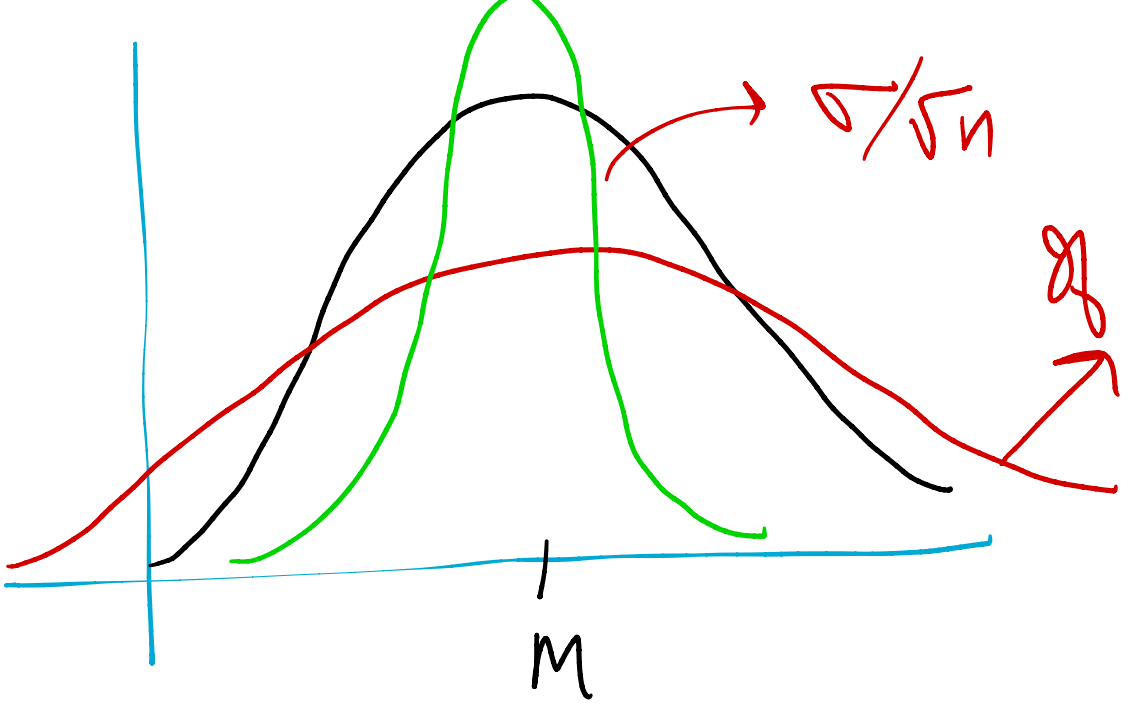
Z, test (or)

t test

Season (1, 2, 3, 4)




What's her job?
She's a statistician
The p-value is the probability that H0 is true, right?
Yes honey, that's close enough
Your foster parents are dead

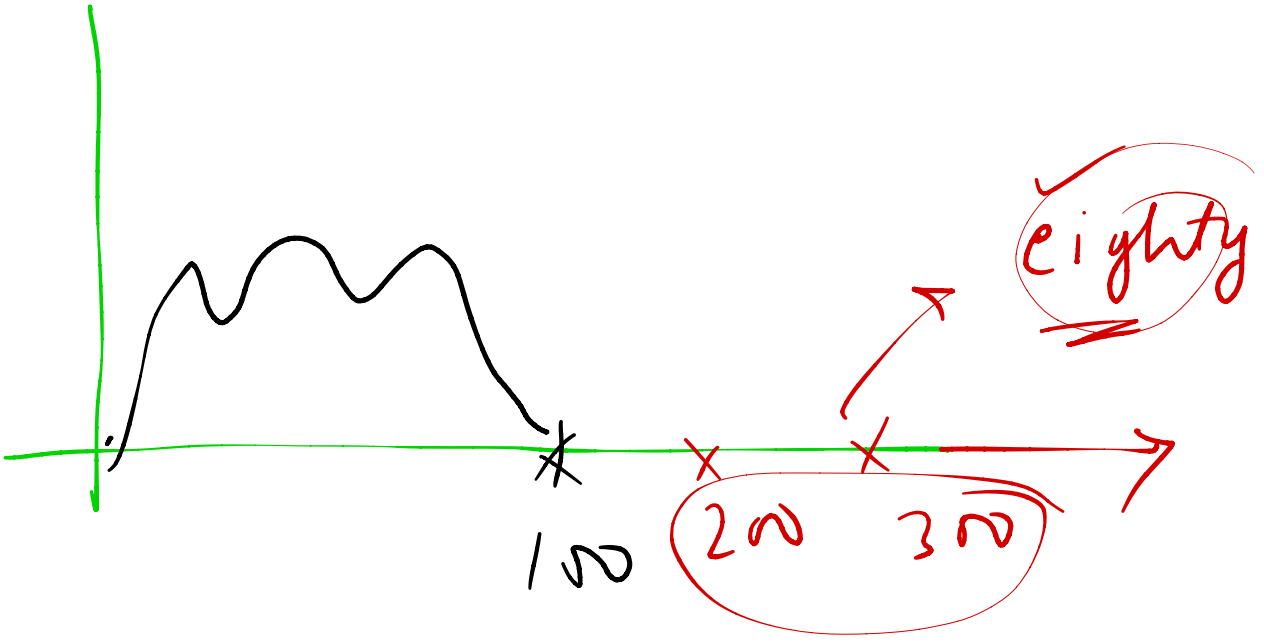


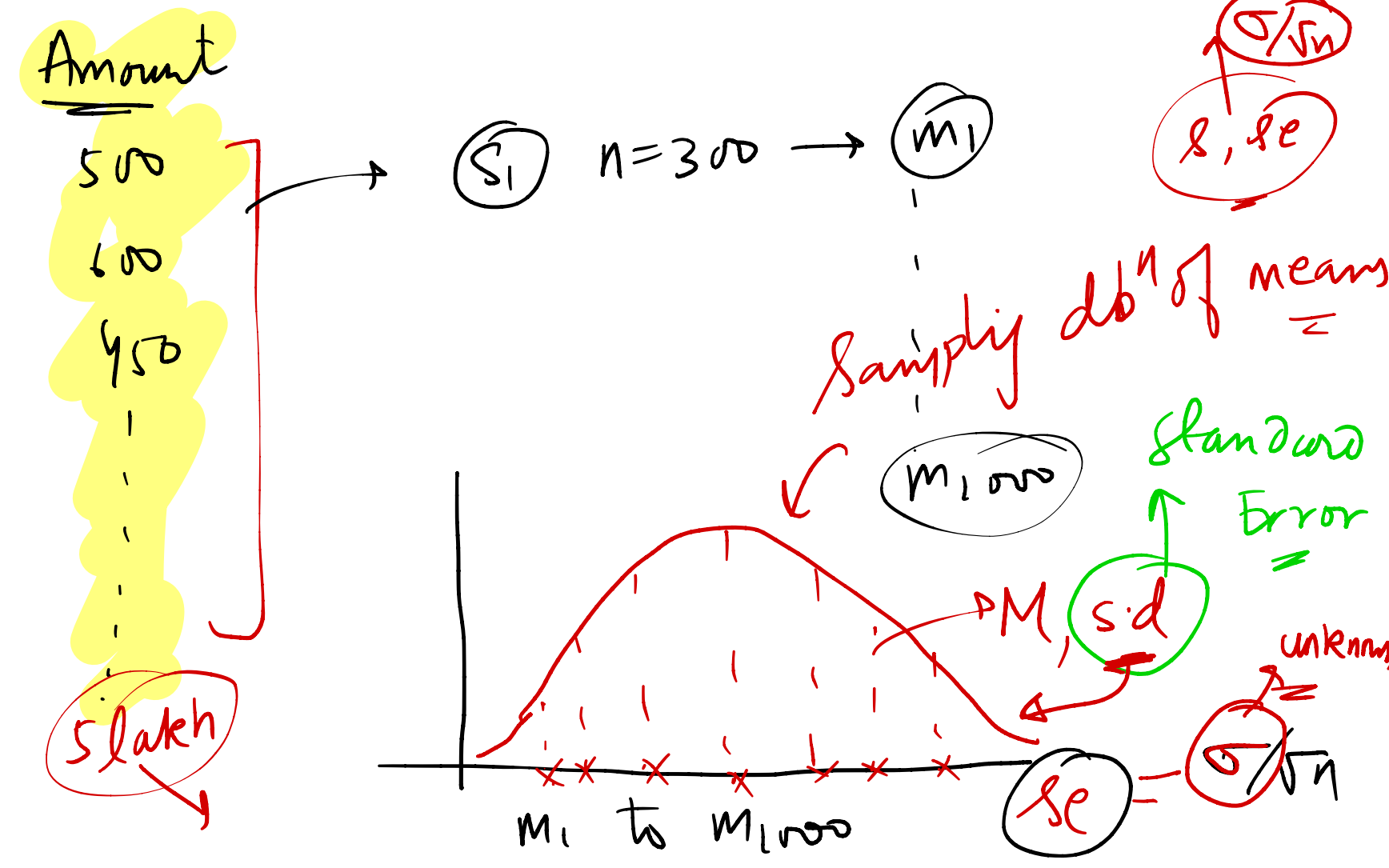
If n is small

↓

fat / narrow

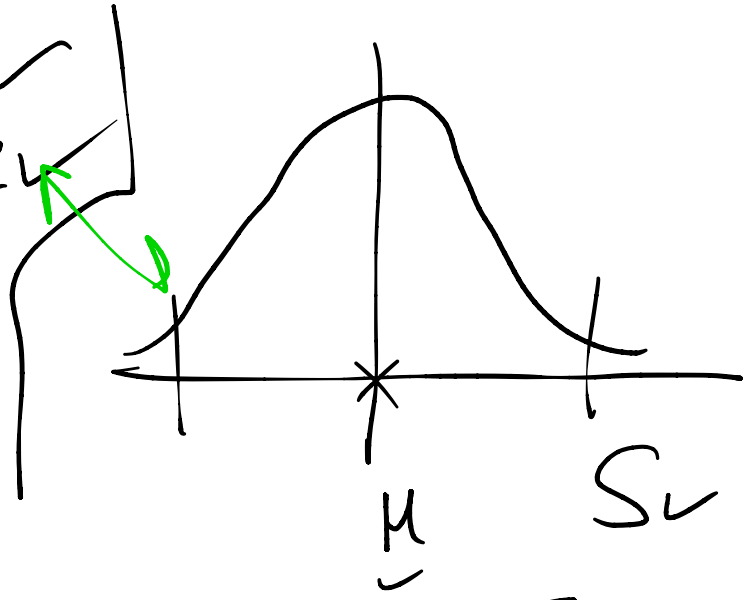
If n is large.


eighty
100
200
300


Amount
500
600
450
5 lakh
S1
n=300
m1
σ/√n
s, se
Sample dbn of means
m1000
Standard
Error
M, s.d
se
m1 to m1000

Computed $\bar{x}$ and SE

$[\bar{x} \pm 1.96(SE)]$



$[\mu \pm 1.96\,S]$