→ Hypothesis Testing

↓

Belief

↙ ↘

Historic — New —

Stats → (Approximation)

(Not Strict)

$\bar{x} = 6$

Interval Estimate

10 min

08:00 Am → 08:10 Am

On-time

Early

07:51 Am

late

08:11 Am

8:01

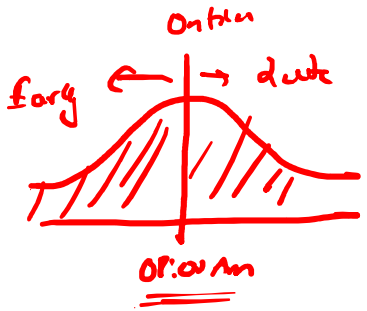
6 2 - 10

3 = 4 = 5 = 6

Before ← Covid → After

Avg Salary of Data Scientist

Avg Salary of Data Scientist



07:59 Am

Early

Estimate

Point ↙ ↘ Interval

$P_1 = P_2$ (Exactly)

(Range)

$LL < x < UL$

(=)

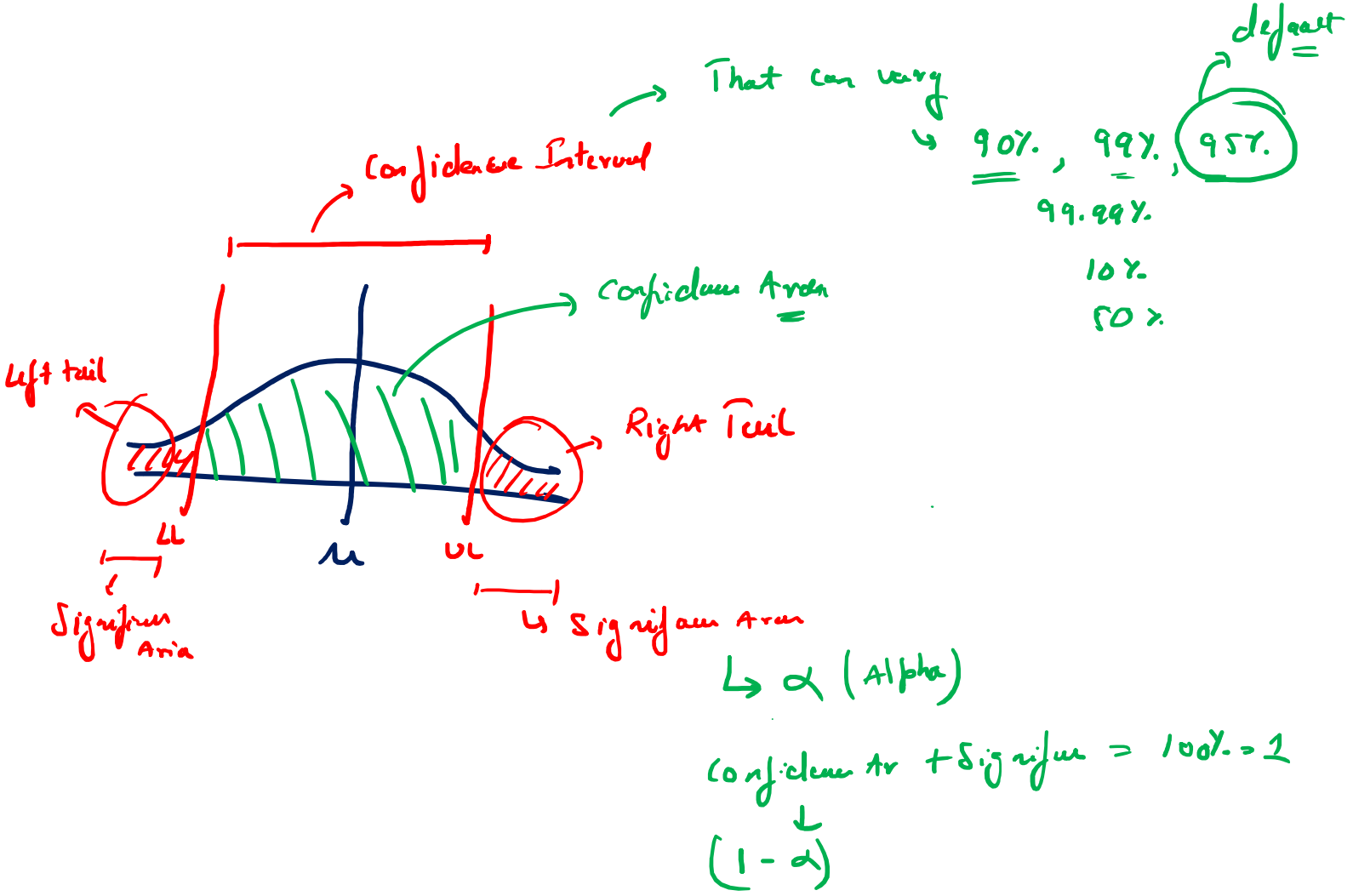
→ 8:00 AM → date

↓ 08:01 Am

on-time

Point Estimate = (Mean) $\bar{S} \neq 6$

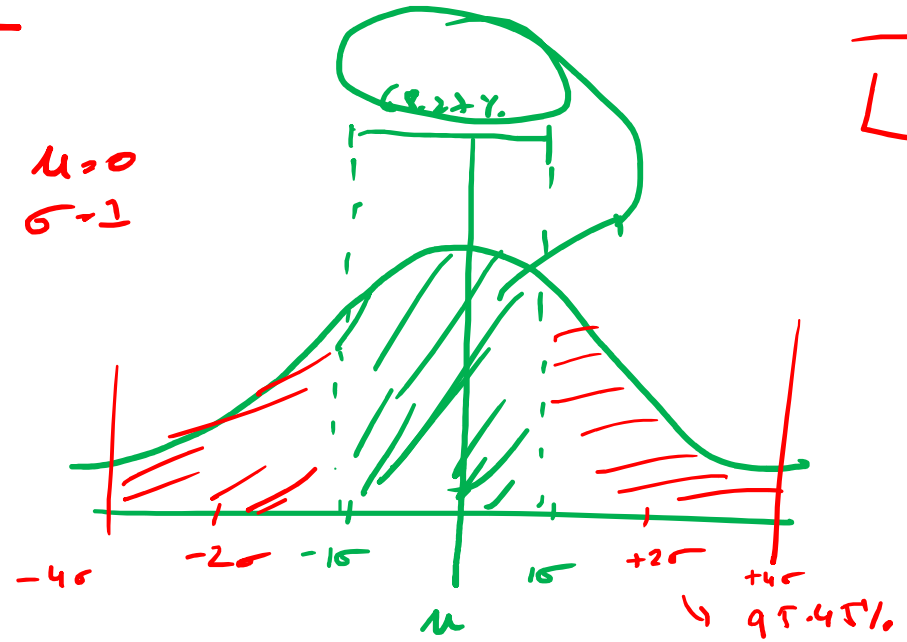
(Strict)


Confidence Interval
That can vary
90%, 99%, 95%
default
99.99%
10%
50%
Confidence Area
Left tail
Right Tail
LL
μ
UL
Significance Area
Significance Area
↳ α (Alpha)
Confidence + Significance = 100% = 1
(1 - α)

Empirical formula

$$AUC = \int_{-1}^{+1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

SND: $\mu = 0$, $\sigma = 1$



$\mu \pm 4\sigma \Rightarrow 99.99\%$

Toothpaste

Same

Considered to be equal

Avg. Salary of a Data Scientist → Survey → population → forbes → survey → current data

you → Aug 2023 → collect data → your company

$\mu = \bar{x}$

Historical Belief = No Impact

challenge has Impact $\mu \neq \bar{x}$

Before ← Covid → After

$\mu = \$ 100{,}000$ → 1000 → 7500. 125000

$\mu = \$ 90{,}000$ → 100

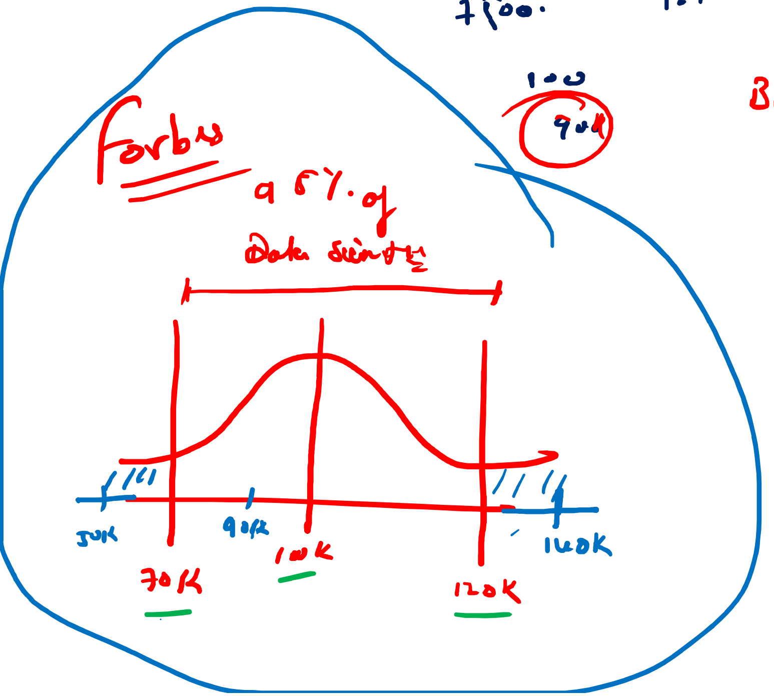
→ 2024

2023 Report + 2022 data

Point Estimate

Because of Avg. point Estimate will not make sense

65K
120K

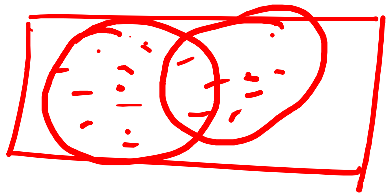


100
90K

$\mu = \$ 140{,}000$

$\mu = \$ 50{,}000$

significantly diff...



Sample size → Population size

pre 2021 272

Pop ✗ Sample

M T W Th F Sat Sun

250 300 200 280 340 270 310

200 — 340

↓

Normal

$\mu = 270g$

Eats food ??

Rice / Chapati

Amount

100g

1000g

↓

winner

↓

2 cm

3 cm

7 values → 7 days (10 cm)

After

1 - Months

→ 2 days (weekend)

$\bar{x} = 290g$

$\bar{x} = 400g$

$\bar{x} = 100g$

→ Frame the hypothesis

→ figure out the μ, σ, CI, $\bar{x}$

→ compare if $\bar{x}$ lies inside CI, outside CI

↓ Historic ✓

↓ History ✗ → process has changed

95%

99.99%

50%?

→ 60%

99.45%

10K

Fitness center → workout

Notion

7 days → 5 kgs → 1 kg → 95.45%

95.45%

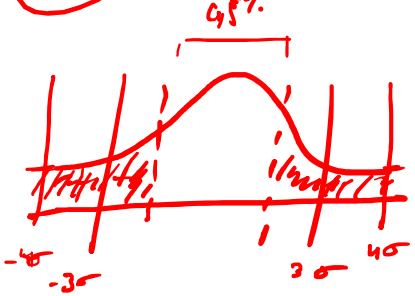
$\mu \pm 2\sigma$ → 99.99%

5 kgs

$5 - 2\times1 = 3$ kg

$5 + 2\times1 = 7$ kg

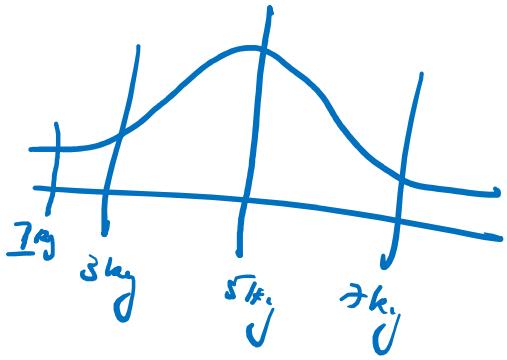
95%



Survey → 5 people

	D_0	D_1	→	weight loss $D_1 - D_0$
1				
2				
3				
4				
5				

→ 1 kg



$H_0 \Rightarrow \mu = 5$ kg → Null hypothesis

$H_A \Rightarrow \mu \neq 5$ kg → Alternate hyp.

→ Framing the Hypothesis