Confidence Interval

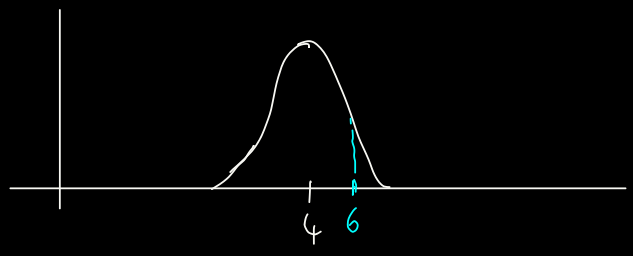
- Recap Gaussian
- Recap CLT
- Confidence Interval using CLT
- Bootstrapping

From a few samples ⟶ conclusions about population

Eg ① Exit poll
② Drug recovery time

} range
How to decide this

The average time taken for customers to complete a purchase is 4 minutes with a standard deviation of 1 minute. Find the probability that a randomly selected customer will complete a purchase within 6 minutes? Assume Gaussian



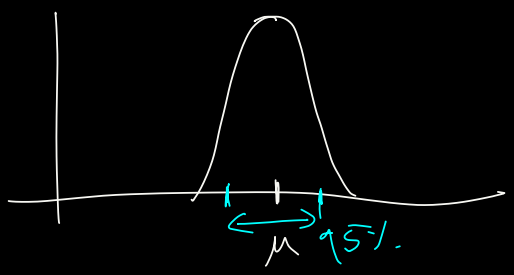
$$z = \frac{6-4}{1}$$

$$\text{norm.cdf}(3)$$

Height: 100 samples → sample mean = 65"
→ Std dev = 2.5

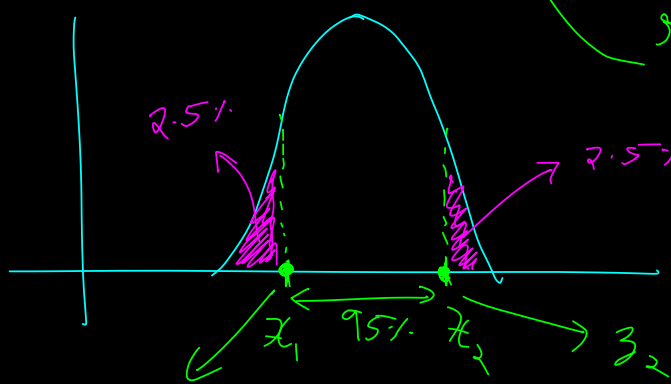
Sample mean distribution

$n = 100$



Stddev: $\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{\sqrt{100}}$

95% Confidence Interval (Typical values will be 95 or 99 rarely 90)



$z_2 = \text{norm.ppf}(0.975)$

$z_1 = \text{norm.ppf}(0.025)$ z_1

Std error $= \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{100}}$

$x_1 = 64.5$ $x_2 = 65.4$

95% CI $[64.5, 65.4]$ → You expect the pop mean to lie here 95% of the time

Bootstrapping

SDE-2 Salary (Survey 1) → 35L, 36L, 33L, 37L, 34L, 35L

avg = 35L

CI (34, 36)

SDE-2 (Survey 2) → 20L, 37L, 17L, 50L, 53L, 35L

avg = 35L

Survey 1 would have better confidence (24, 46)