Central Limit Theorem

From samples, what can we infer about population

① Central Limit Theorem (today)

② Confidence Interval (next class)

1) Height $\longrightarrow$ avg $\simeq 66.3$ S tdev $= 3.84$

① 5 samples at a time (randomly)

- avg of these 5 people
  $\longrightarrow$ around 66

- sample again (random)
  $\longrightarrow$ around 66, but a different number

Is sample mean a random variable? Yes!

1000 such sample means are given of 5 samples

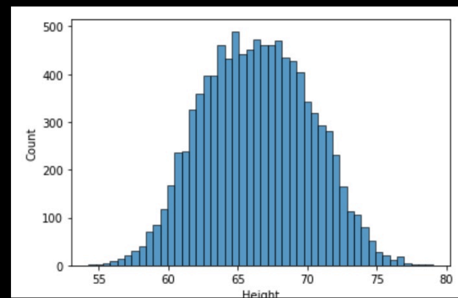
what is the central tendency of these 1000 numbers

very close to 66.3 (population mean)

population

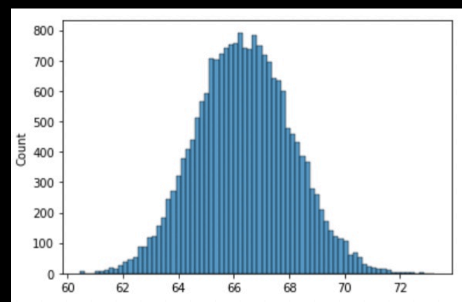
mean → 66.36

std dev → 3.84



(Samples of size 5) 20,000 of such

observation: std dev is lesser than pop std dev



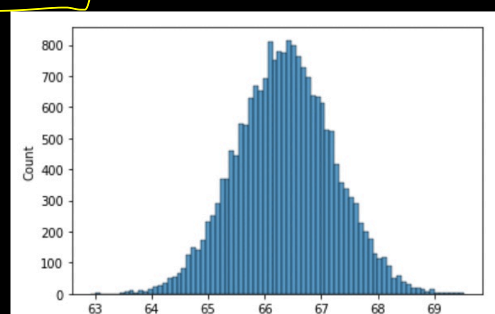
mean ≃ 66.3

mean of sample mean ≃ pop. mean

1

std dev = 1.7

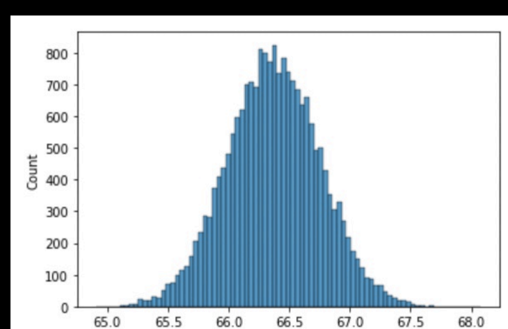
(Samples of size 20) 20,000 such



mean of sample mean ≃ pop. mean

Std dev ≃ 0.85

(Samples of size 100) 20K such



mean of sample means ≃ pop mean

mean ≃ 66.36

Std dev = 0.38

If "n" denotes the sample size and "σ" denotes the population std dev then the std dev of the sample mean

$$\boxed{\text{Standar Error} = \frac{\sigma}{\sqrt{n}}}$$

We saw

$n = 5$

$n = 20$

$n = 100$

$\sigma = 3.84$

Standard error is the std dev of the sample mean

Let $\bar{X} = \dfrac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$

denote the sample mean

Then $\bar{X}$ follow a Gaussian distribution

with $E[\bar{X}] = \mu$ pop. mean

Std dev of $\bar{X} = \dfrac{\sigma}{\sqrt{n}}$

$\bar{X} \sim N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$

X_i is a random variable from same population

μ: pop mean

σ: pop. std dev

If n is large ($n > 30$)

σ is finite

Original distribution need not be Gaussian

Systolic blood pressure of a group of people is known to have an average of 122 mmHg and a standard deviation of 10 mmHg

Calculate the probability that the average blood pressure of 16 people will be greater than 125 mmHg.

What is n? $n = 16$

$\mu = 122$

$\sigma = 10$

$$\text{Std error} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{16}} = 2.5$$

Draw distribution of sample mean

$$z = \frac{125 - 122}{2.5} = 1.2$$

Std dev = 2.5
($n = 16$)

prob = 1 − norm. cdf (z)

Weekly toothpaste sales have a mean 1000 and std dev 200. What is the probability that the average weekly sales next month is more than 1110?

$\mu = 1000$

$\sigma = 200$

$n = 4$

$\text{Std error} = \frac{200}{\sqrt{4}} = 100$

Sample mean
$n = 4$

Std dev = 100

$z = \frac{1110 - 1000}{100} = 1.1$

$1 - \text{norm.cdf}(z) = 0.13$

In an e-commerce website, the average purchase amount per customer is $80 with a standard deviation of $15. If we randomly select a sample of 50 customers, what is the probability that the average purchase amount in the sample will be less than $75?

$n = 50$

$\text{Std error} = \frac{15}{\sqrt{50}} = 2.12$

$z = \frac{75 - 80}{2.12} = -2.34$

$\text{norm.cdf}(z) = 0.009$

Sampling