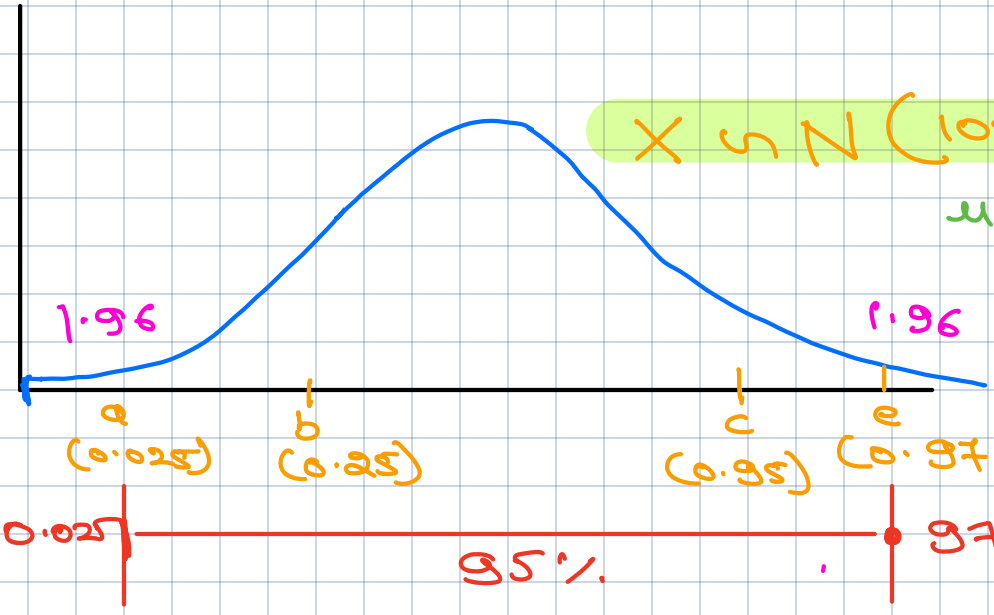
Welcome Quiz Given Normal Distribution



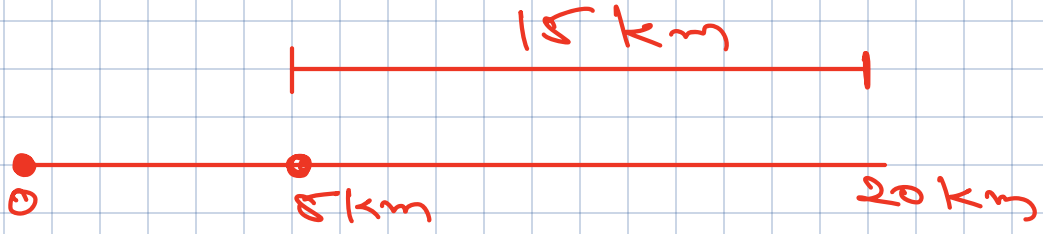
Q: Which of the following options pairs covers 95% of Total Data

(a) (a, b)

(b) (a, c)

(c) (b, c)

(d) (a, e) ✓



Follow up:

Q: Can you find Exact Values for your Answer?

a → ?

e → ?

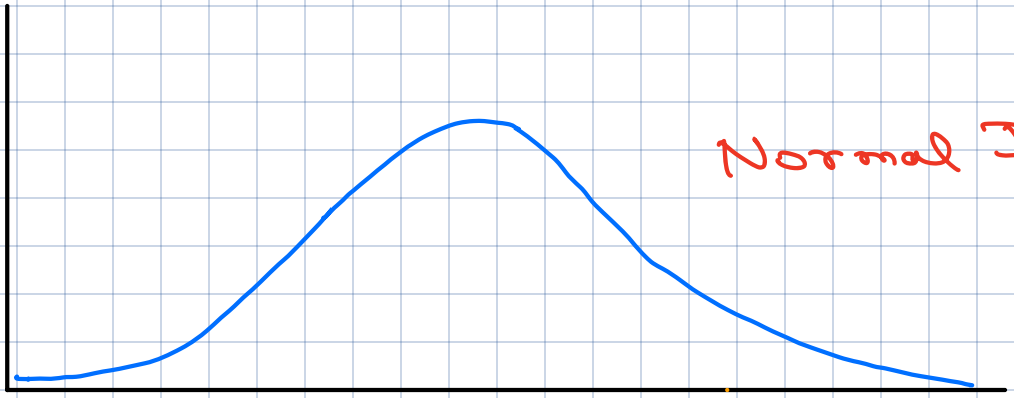
Agenda

- CLT
- Confidence Intervals

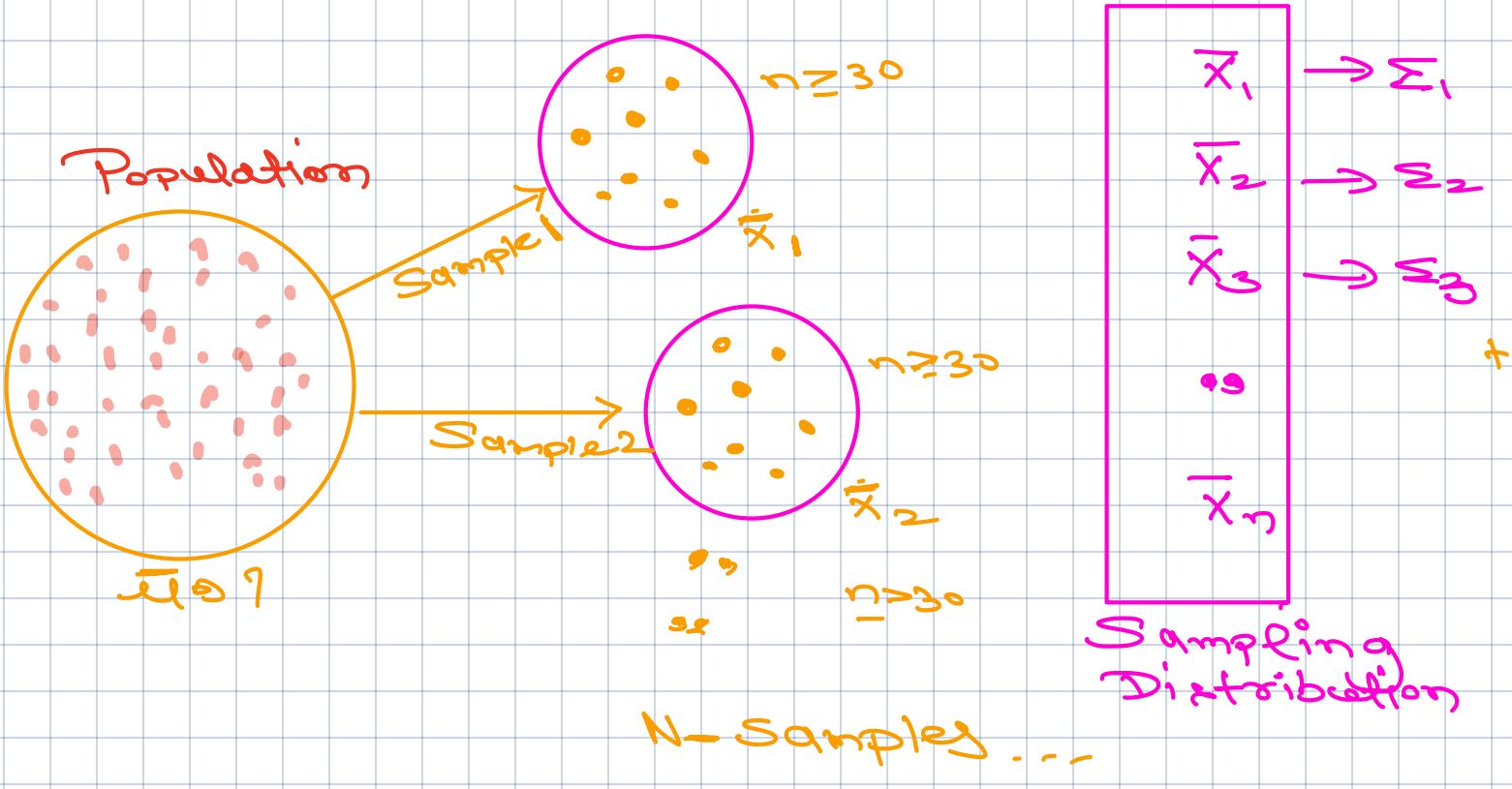
Central Limit Theorem

Remember Sampling Distribution?

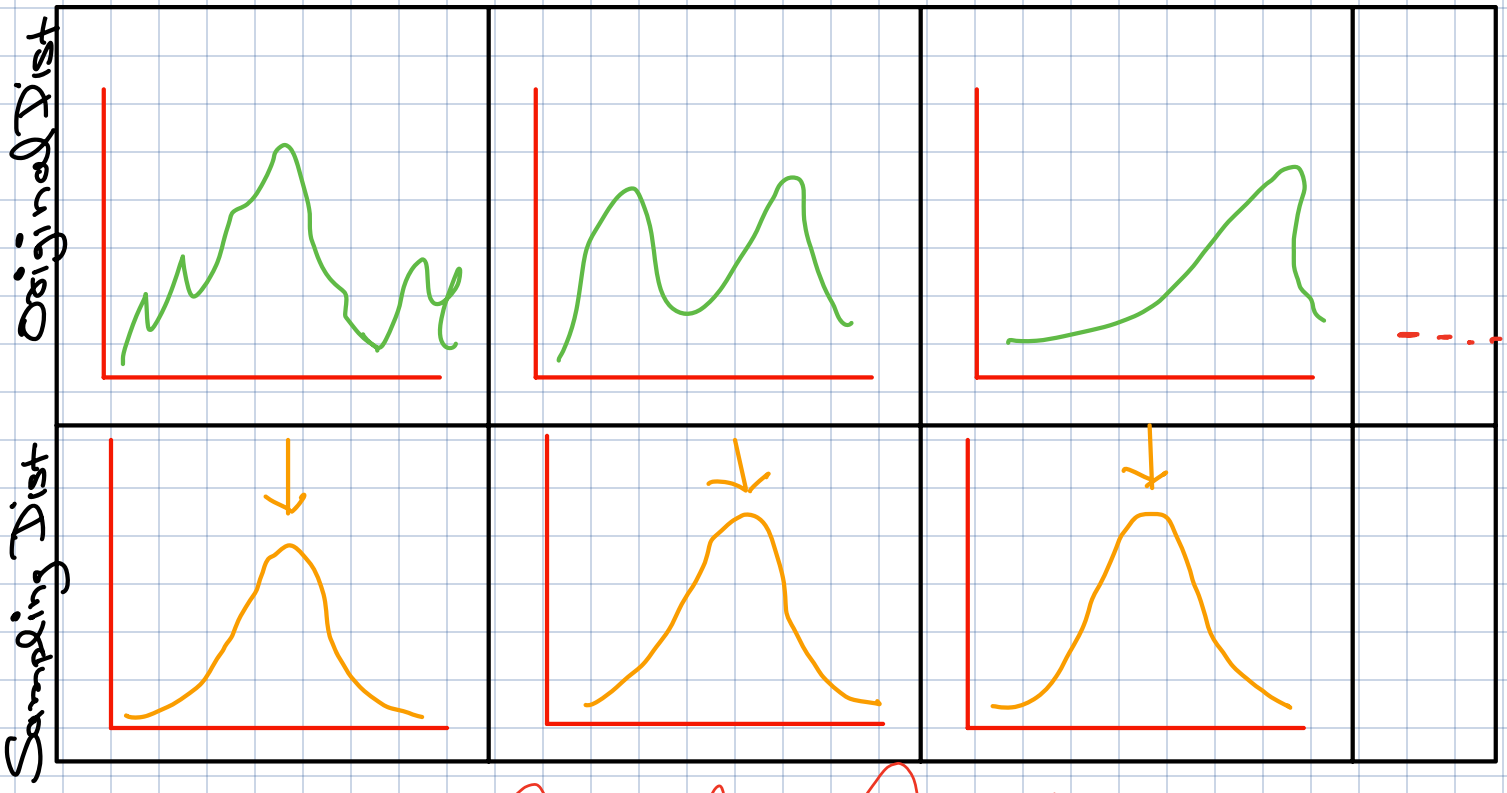
Collection of means from n-samples

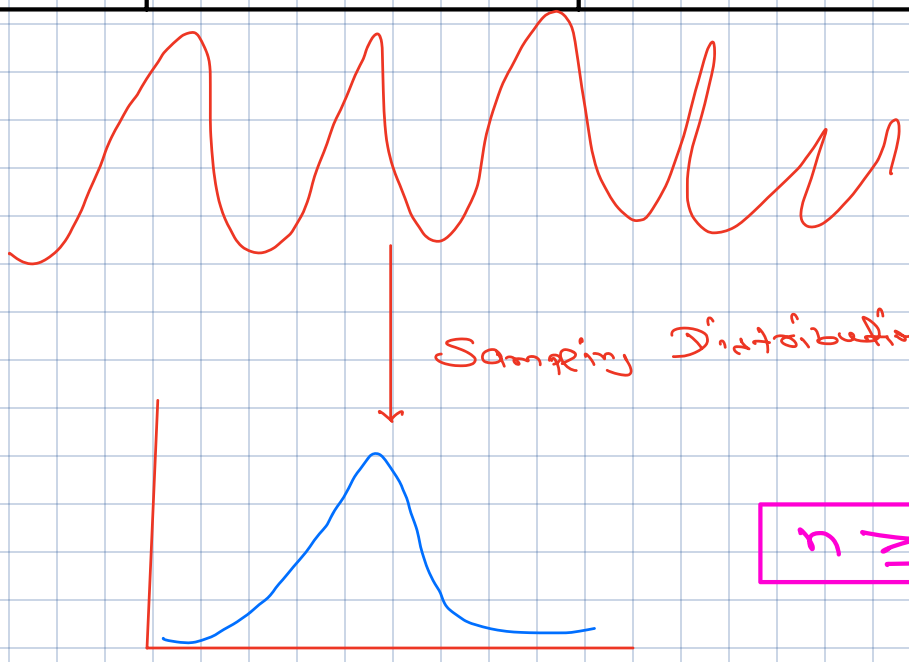


Normal Distribution



Goal: Estimate μ by MLE

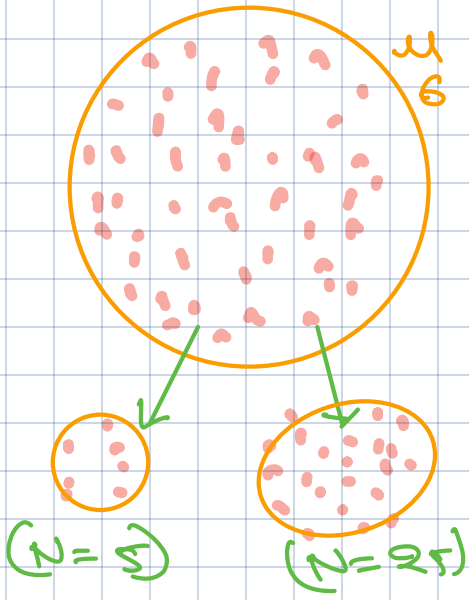


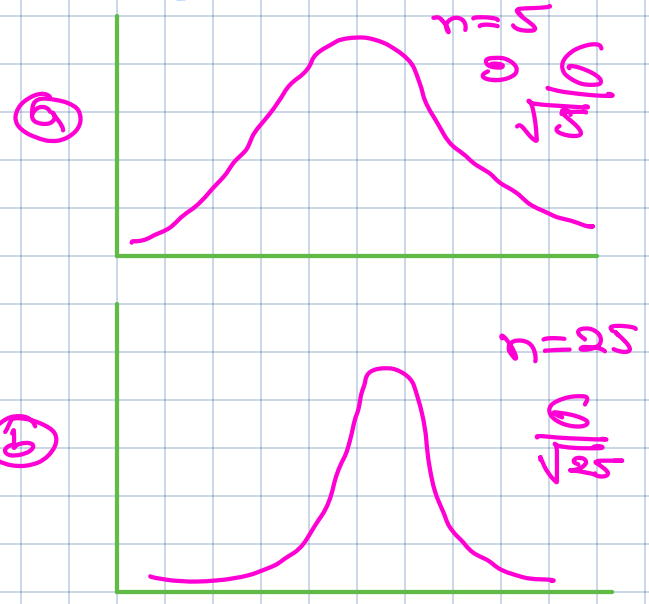


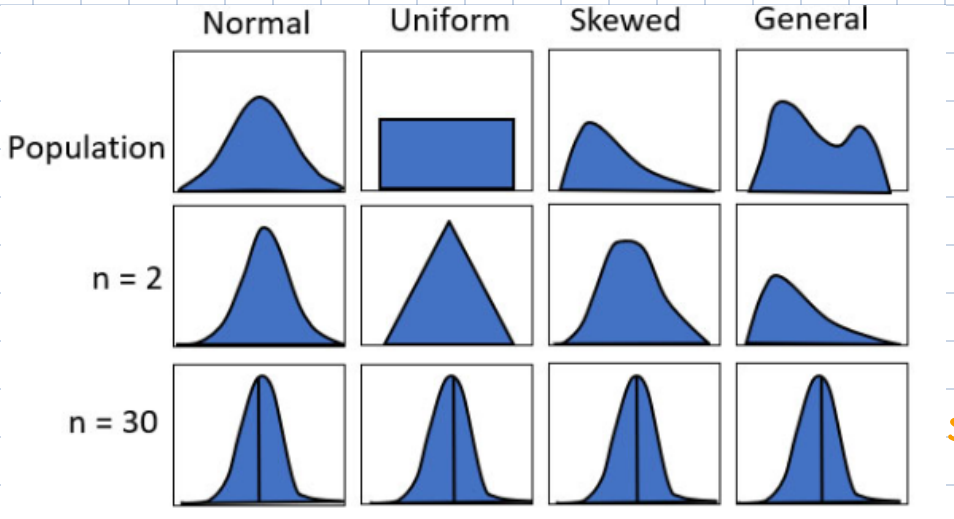
$n \geq 30$

CLT : As the sample size increases The sampling Distribution becomes almost Normal regardless of population Distribution

① Sample Size and Normality







Sampling Dist

Sampling Dist

② Sample Size and Standard Deviation

$SE = \frac{\sigma}{\sqrt{n}}$ or $\frac{s_1}{\sqrt{n}}$ $\quad SE \propto \frac{1}{\sqrt{n}}$

↓

This is SD of Sampling Dist

Larger the sample size, Lower the Standard Deviation of Sampling Distribution (SE)

Conclusion

① Mean of sampling Distribution gets closer to population mean as sample size increase

$$\mu = \mu_{\bar{x}_{20}} = \mu_{\bar{x}_5}$$

② As we increased Sample size the SD of sampling Distribution (SE)

$$\sigma > \sigma_5 > \sigma_{20}$$

Sampling Distributions

$$SE(\sigma_{\bar{x}}) = \boxed{\frac{\sigma}{\sqrt{n}}} \text{ or } \frac{s}{\sqrt{n}}$$

Conditions for CLT

1) **Randomization:**

Data should be randomly Sample with every instance Raving Equal Chance

2) **Independence:**

Every sample Should be independent and should not impact Others

3) **Large Sample Size:**

Larger the sample size, reliable the estimates.

↓

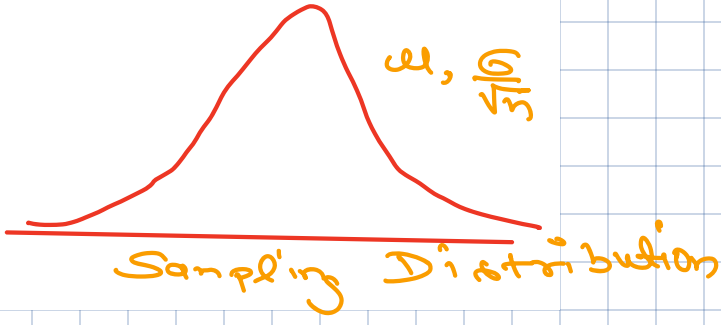
General Rule of thumb $n \geq 30$

Summazing CLT

$$\bar{X} \backsim N(\mu, \frac{\sigma}{\sqrt{n}})$$ ⇒ Sampling Distribution

Where:

- $\bar{X}$ is the sampling distribution of the sample means
- $\backsim$ means "follows the distribution"
- N is the normal distribution
- μ is the mean of the population
- $\sigma / \sqrt{n}$ is the standard error



Questions

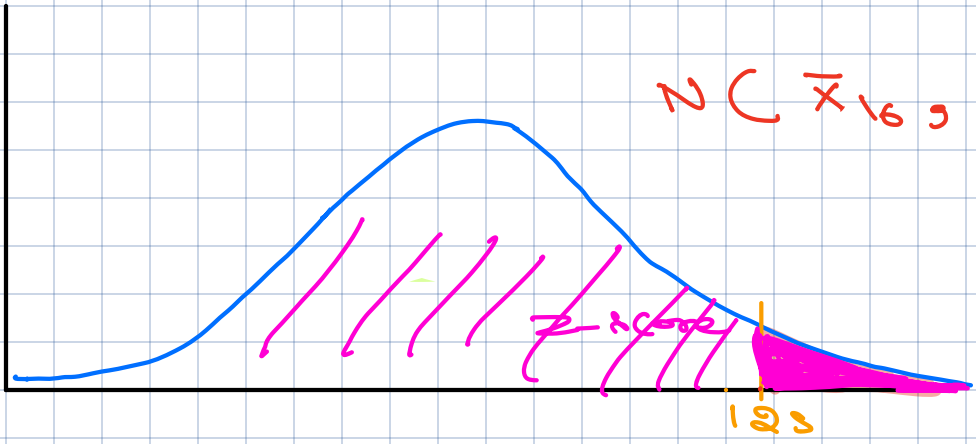
Systolic blood pressure of a group of people is known to have an average of 122 mmHg and a standard deviation of 10 mmHg.
Calculate the probability that the average blood pressure of 16 people will be greater than 125 mmHg.

Answers

$\mu = 122$

$\sigma = 10$

$P(\bar{X}_{16} > 125)$



$N\left(\bar{x}_{16}, \frac{\sigma}{\sqrt{16}}\right)$

Avg Blood pressure of samples of size 16

$\bar{x}_{16} \Rightarrow \mu = 122$

$SE = \frac{\sigma}{\sqrt{16}} = \frac{10}{\sqrt{16}}$

$Z\text{-score} = \frac{125 - 122}{SE}$

$cdf(Z\text{-score}) \Rightarrow P(X \leq 125)$

$Ans = 1 - P(X \leq 125)$

Questions

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?

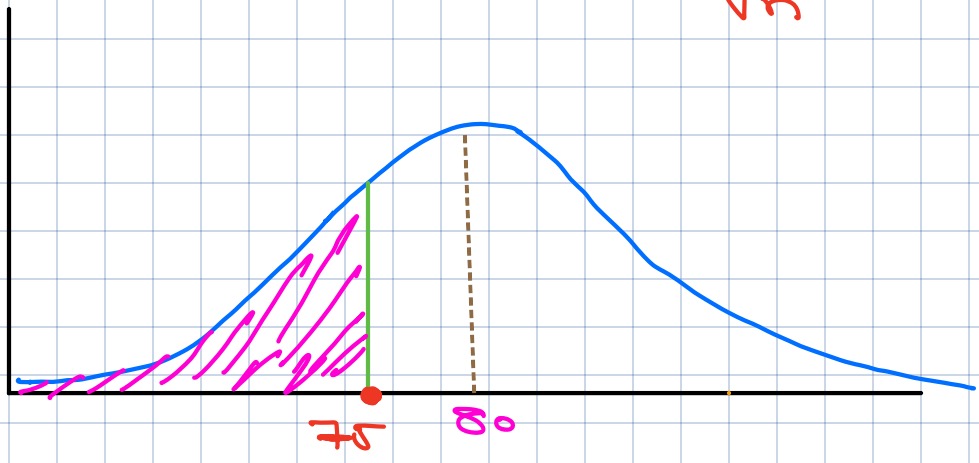
Answers:

$\mu = 80$

$\sigma = 15$

$\mu_{\bar{x}} = 80$

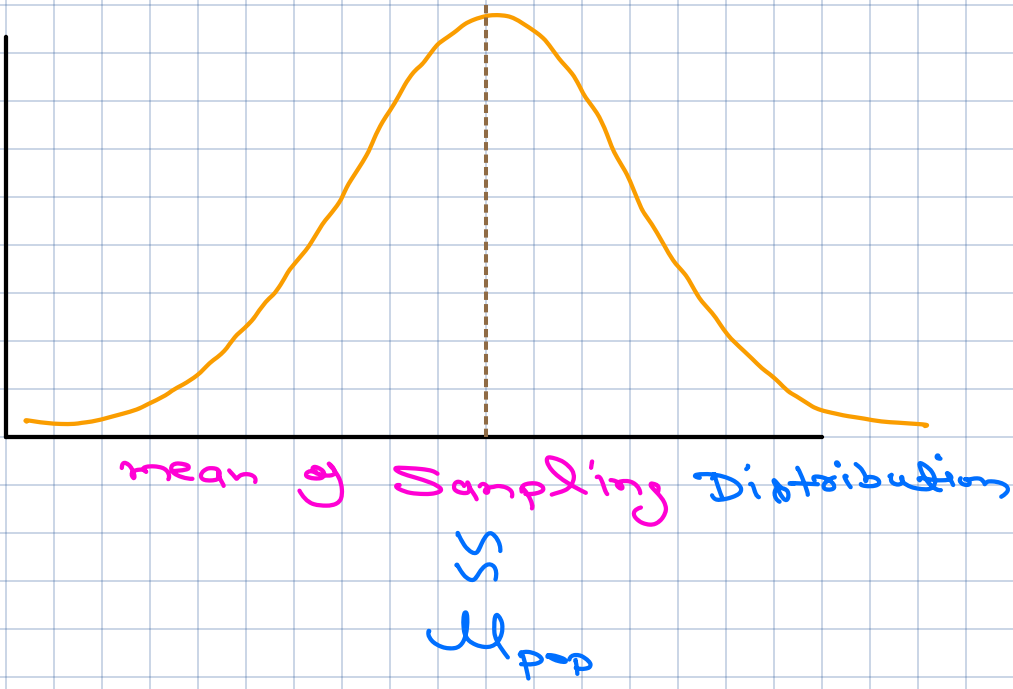
$SE = \frac{\sigma}{\sqrt{n}}$



$P(\bar{x}_{50} \leq 75)$

Confidence Intervals (CI)

Goal: Salary Estimation of MLE



Idea: Instead of giving a single value as estimate, can we give a range of values and assign Confidence Level

Point Estimate: 32 L

Range Estimate: 31 — 33L (95%)

CL: 95% of the times the Avg delay will be between 31 and 33 LaKhs

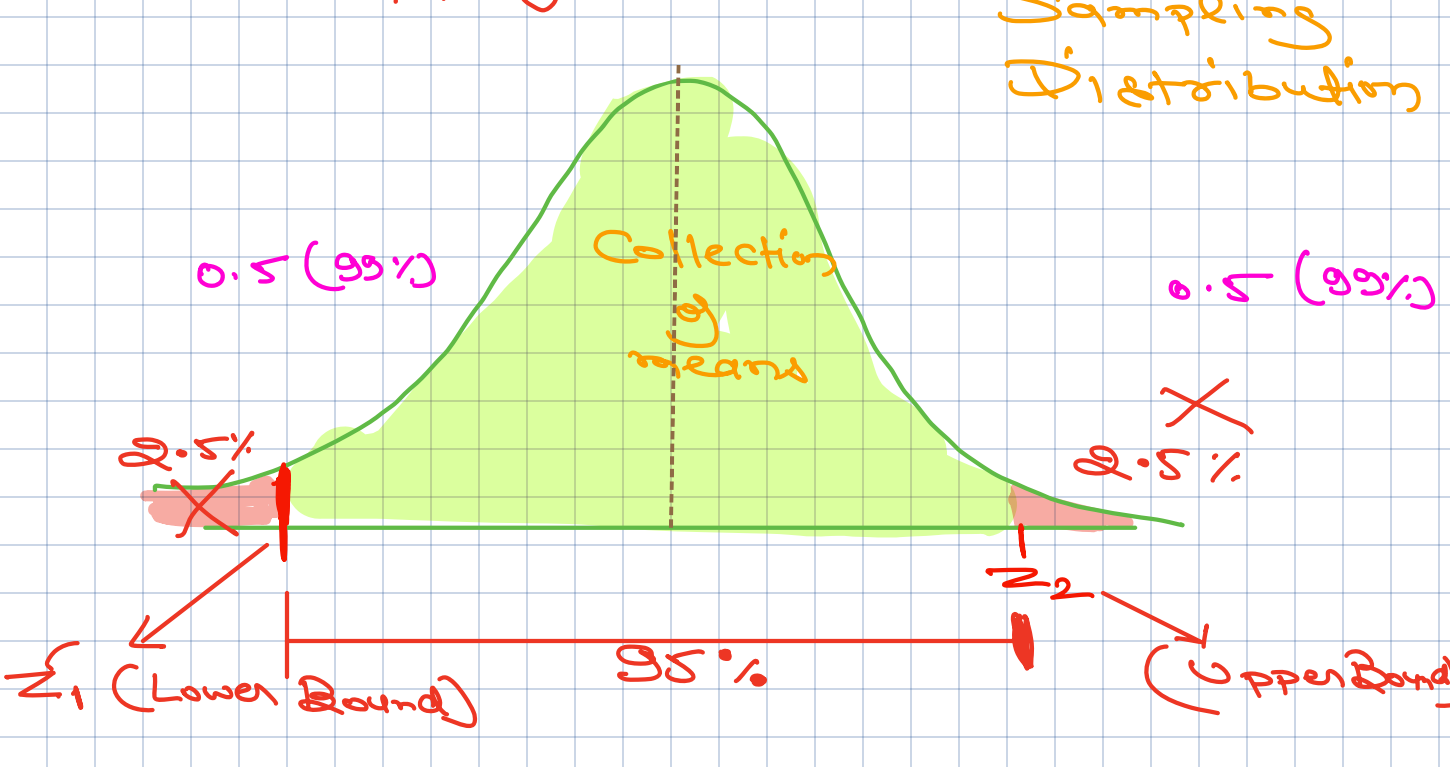
Confidence Intervals

[Lower Bound , Upper Bound]

associated Confidence Level

Methods of Calculating CI

1. CLT
2. Bootstrapping



z_1 = norm. ppf (0.025)

z_2 = norm. ppf (1 − 0.025)

or

0.975

$z = \frac{x_1 - \mu}{SE(\bar{x})} = \mu - z \times \frac{\sigma}{\sqrt{n}}$

$$\mu - z \times \frac{\sigma}{\sqrt{n}} \quad , \quad \mu + z \times \frac{\sigma}{\sqrt{n}}$$

$$CI = \mu \pm z_{1-\alpha} * \frac{\sigma}{\sqrt{n}}$$

95% → z_1 = 0.025 → , z_2 = 0.975 →

99% → z_1, z_2 = ±

Questions

The mean height of a sample of 100 adults was found to be 65 inches, with a standard deviation of 2.5 inches.
Compute 95% confidence interval

Answers

$\bar{x}_{100} = 65$

$s = 2.5$

$SE = 2.5/\sqrt{100}$

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

```
[3] z_u = norm.ppf(1 - 0.025)
    z_u

    np.float64(1.959963984540054)

[4] x_upper = 65 + z_u * std_error
    x_upper

    np.float64(65.48999099613502)

[5] x_upper = 65 - z_u * std_error
    x_upper

    np.float64(64.51000900386498)
```

$[64.51, 65.489]$

95% CL

Questions

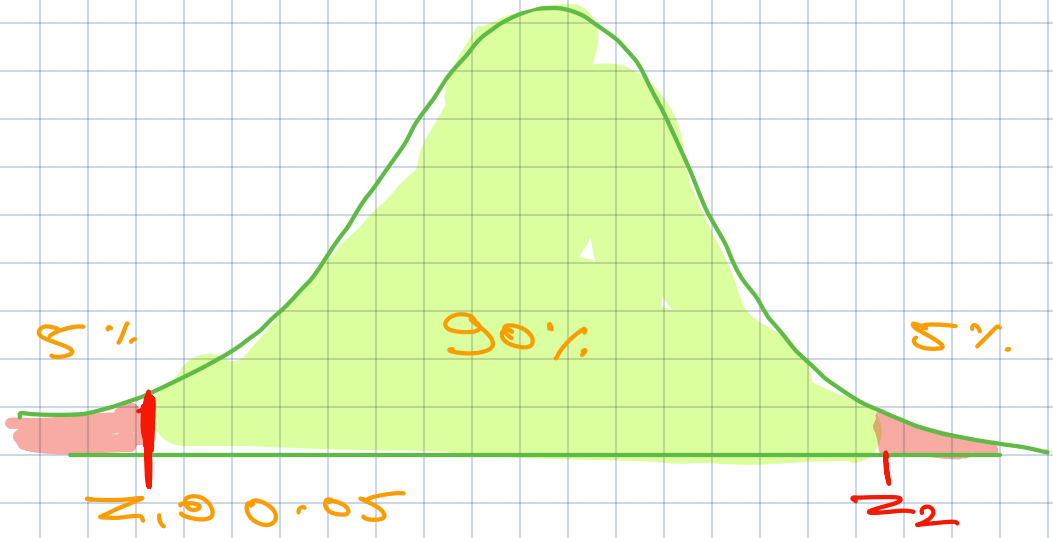
The sample mean recovery time of 100 patients after taking a drug was seen to be 10.5 days with a standard deviation of 2 days
Find the 95% confidence interval of the true mean.

Answers

Questions

From a sample of 80 endangered birds, the average wingspan was found to be 45 cm, with a population standard deviation of 10 cm. What is the correct confidence interval of the mean wingspan of the entire population with 90% confidence.

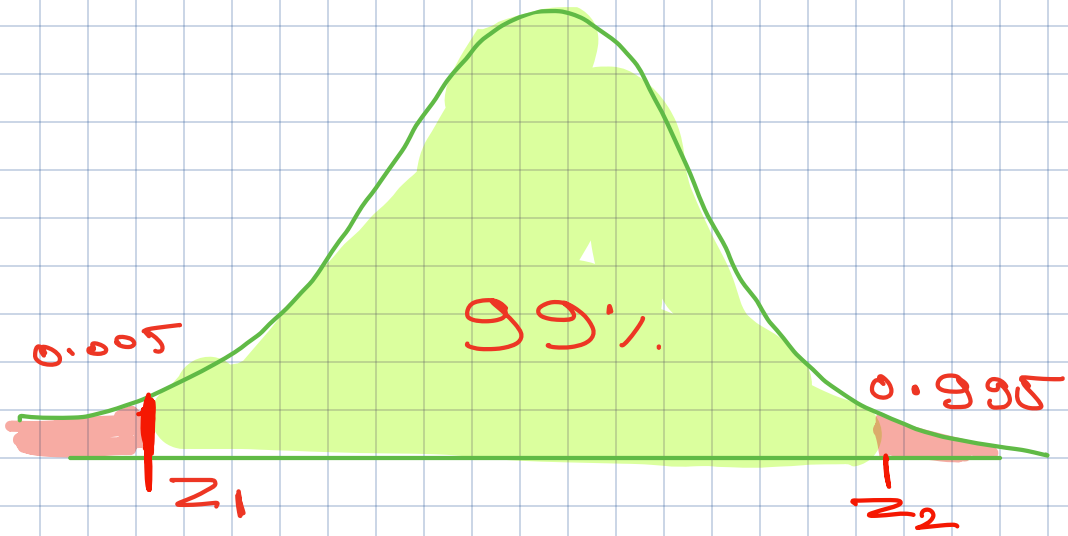
Answers



Questions

In a software project, the team estimates bug resolution time at an average of 6 hours with a standard deviation of 2 hours.
To estimate the mean resolution time with 99% confidence, the project manager samples 25 resolved bugs.
What is the correct confidence interval?

Answers



1 g Error Margin

② CI using BootStrap Method

Step 1: Sample with Replacement

Step 2: Calculate mean and SD of Each Sample to form Sampling Dist of Target Statistics

Step 3: Calculate 2.5 and 97.5 percentile of Sampling Dist

When to Use Bootstrapped Technique?

[20, 25, 37, 17, 50]

↓ BootStrap Sampling

[20, 25, 20, 37, 17]

Why Perform Bootstrap Sampling

① You have no other option but a few samples

② Estimate Median / Mode

Repeat All the Question

Solve with pen and paper