PL 08 : CONFIDENCE INTERVAL ESTIMATION

Video 1: Inferential Statistics : CONFIDENCE INTERVALS (σ Known)

- When estimating a population parameter using a sample statistic, it is never going to be perfect.

- That error or uncertainty, ~~using~~ can be expressed using an interval estimate.

  Point Estimate $\pm$ Margin of error

- Recruit :- to find Standard error of mean, two things req
  (i) The Population Standard Deviation
  (ii) The Sample size.

- Interpretation.

-- The randomness lies in the elements chosen for the sample : NOT the Population mean.

-- It is the probability of obtaining a representative sample.

-- The proportion of samples, size n, for which our estimate, the sample mean $\bar{x}$, is within a certain distance $\pm$ of the true population, μ.

-- The sample mean is either within $\pm$ interval of the true mean or it is not.

XXXX Error XXXX The confidence interval is not the probability that the population mean lies within the interval.

-- The probability that the sample mean of sample size n is within a certain distance $\pm$ of the true population mean.

- Example:

To estimate the mean amount spent per customer at a shain, data was collected for 75 customers. The population standard deviation is \$4.

(i) At 95% confidence, what is the margin of error?
(ii) If the sample mean is \$20, what is the 95% confidence interval for the population mean (all customers)?

(i) $\bar{x} \pm 1.96\,\sigma_{\bar{x}}$ ← Margin of error

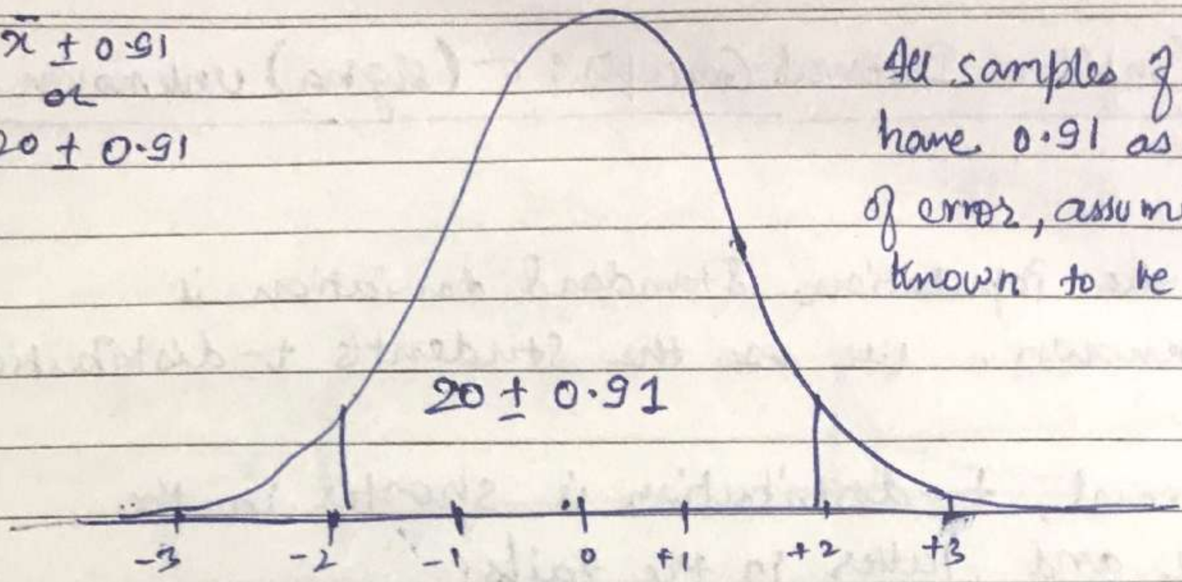
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \Rightarrow \sigma_{\bar{x}} = \frac{4}{\sqrt{75}} \Rightarrow \sigma_{\bar{x}} = 0.46$$

$\therefore$ Margin of error $= \bar{x} \pm 1.96(0.46)$
$= \bar{x} \pm 0.91$

(ii) If sample mean is \$20, what is 95% confidence interval?

$\bar{x} \pm 0.91 = 20 \pm 0.91$

$\bar{x} \pm 0.91$
or
20 ± 0.91



All samples of $n = 75$ will have 0.91 as the margin of error, assuming σ is known to be 4.

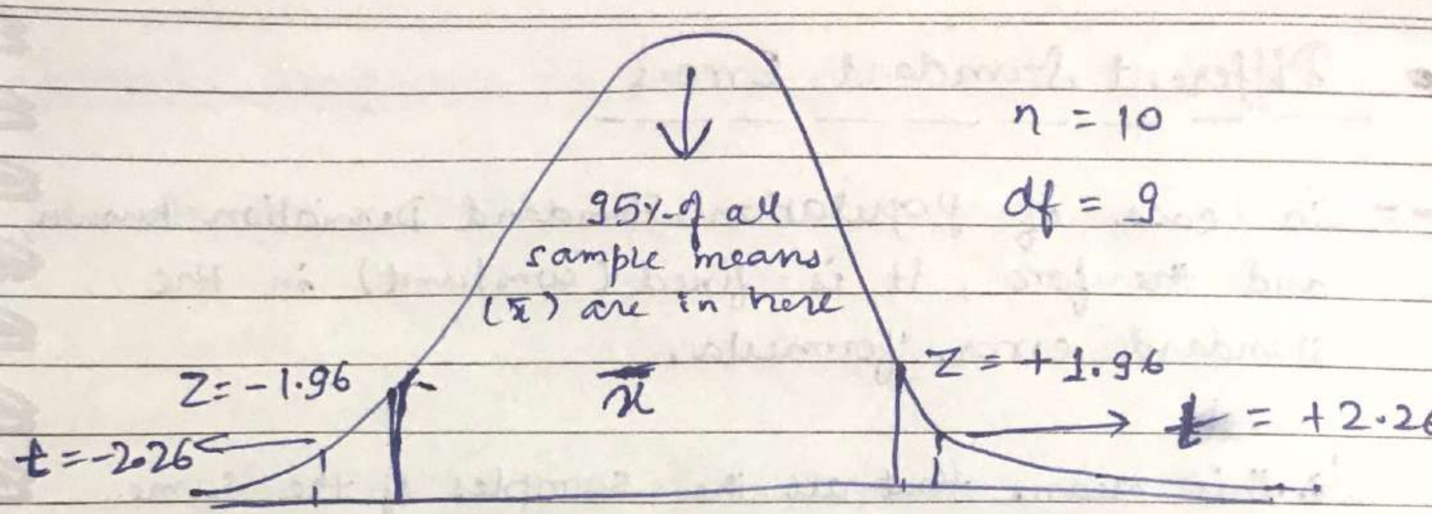
If we take 100 samples of $n = 75$, and make intervals $\bar{x} \pm 0.91$, 95 of them will contain μ.

95% of all intervals made using $\bar{x} \pm 0.91$ will contain the unknown population mean.

"95% confident"

Video 2: Confidence Interval Concepts: σ (sigma) unknown.

- When the population standard deviation is not known, we use the Student's t-distribution.
- In general, t-distribution is shorter in the middle and fatter in the tails.
- More probability in the tails, less near the mean; greater chance of extreme values.
- There is not just one t-distribution.
- There is a t-dist for every sample size.
- Degrees of Freedom $(n-1)$
- Smaller the sample size, the shorter and fatter the distribution; more tail probability
- However as n becomes large, the ~~dis~~ t-dist converges towards the z-dist.



- If we do not know the Population Standard dev σ, we treat the sampling distribution as a t-distribution.

- We assign a t-score to the upper and lower boundary of 95% interval for each sample size.

Degrees of freedom $(n-1)$

→ t-score can be found using t-distribution table.

So, Standard Error Formula :

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

S = Standard Deviation of Sample

- ## Different Standard Errors

-- In case of population Standard Deviation known, and therefore, it is fixed (constant) in the standard error formula.

↳ This means that all the samples of the same size had the same standard error.

-- When σ is unknown, we estimate it with the SAMPLE STANDARD DEVIATION, S

↳ Since every sample will have a UNIQUE, S, samples of the same size DO NOT necessarily have the same standard error.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

$S_{\bar{x}}$ = Standard error of mean
S = Standard deviation of sample

↳ The randomness of sample selection is represented in its standard deviation and therefore its standard error.

↳ Interpretation :- Refer last video notes

Video 3: Confidence Interval Problems.

1. At 95% confidence, what is the margin of error?

$n = 15$ $\bar{x} \pm 2.145 s_{\bar{x}}$

$s = 4$

$df = 14$ and $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4}{\sqrt{15}} = 1.03$

$\to \bar{x} \pm 2.145(1.03)$

$= \bar{x} \pm 2.21$

2. $n = 30$ Given

$\bar{x} = 3661.5$ So, Standard error of mean

$s = 1958$ $\Rightarrow s_{\bar{x}} = \frac{1958}{\sqrt{30}} \Rightarrow s_{\bar{x}} = 357.48$

At 95% Confidence interval, find margin of error?

(i) Use t-distribution, for 95% i.e 0.05 for 2 tails or 0.025 for 1 tail

for df $(n-1)$ i.e. $30 - 1 = 29$

so, t-score found $= 2.045$.

$\therefore$ Margin of error $= \bar{x} \pm 2.045 s_{\bar{x}}$

$= 3661.5 \pm 2.045(357.48)$

↳ Interpretation: Refer last video notes

Video 4 : Inferential Statistics : Estimating Sample Size

- Instead of calculating margin of error, we are going to choose margin of error and try to find/estimate the SAMPLE SIZE.

- : $\boxed{\text{Point Estimate} \pm \text{Margin of error}}$

- $\bar{x} \pm Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \Rightarrow E = Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

-- We need to find n.

So, rearranging the formula

$$E = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{Z_{\alpha/2}\, \sigma}{E}$$

$$n = \frac{(Z_{\alpha/2})^2 \sigma^2}{E^2}$$

Catch :

To find n, we need σ. But in most cases, we do not know σ.

We have following options

(i) Estimate σ from previous studies

(ii) Conduct a pilot study

(iii) Use a judgement or guess (high - low) divided by 4.

INTERPRETATION :

Q) What minimum sample size is necessary to produce 95% confidence that the sample mean is ± 8 of the true population mean ?"

→ A larger sample is more representative of the overall population.

→ So a larger sample size will be required for :-

(i) A smaller margin of error requirement.

(ii) A higher level of confidence.

(iii) Or both.