Topic 5 Confidence Interval Estimation

Tutorial agenda:

1. go through key concepts
2. tutorial questions
3. Q&A

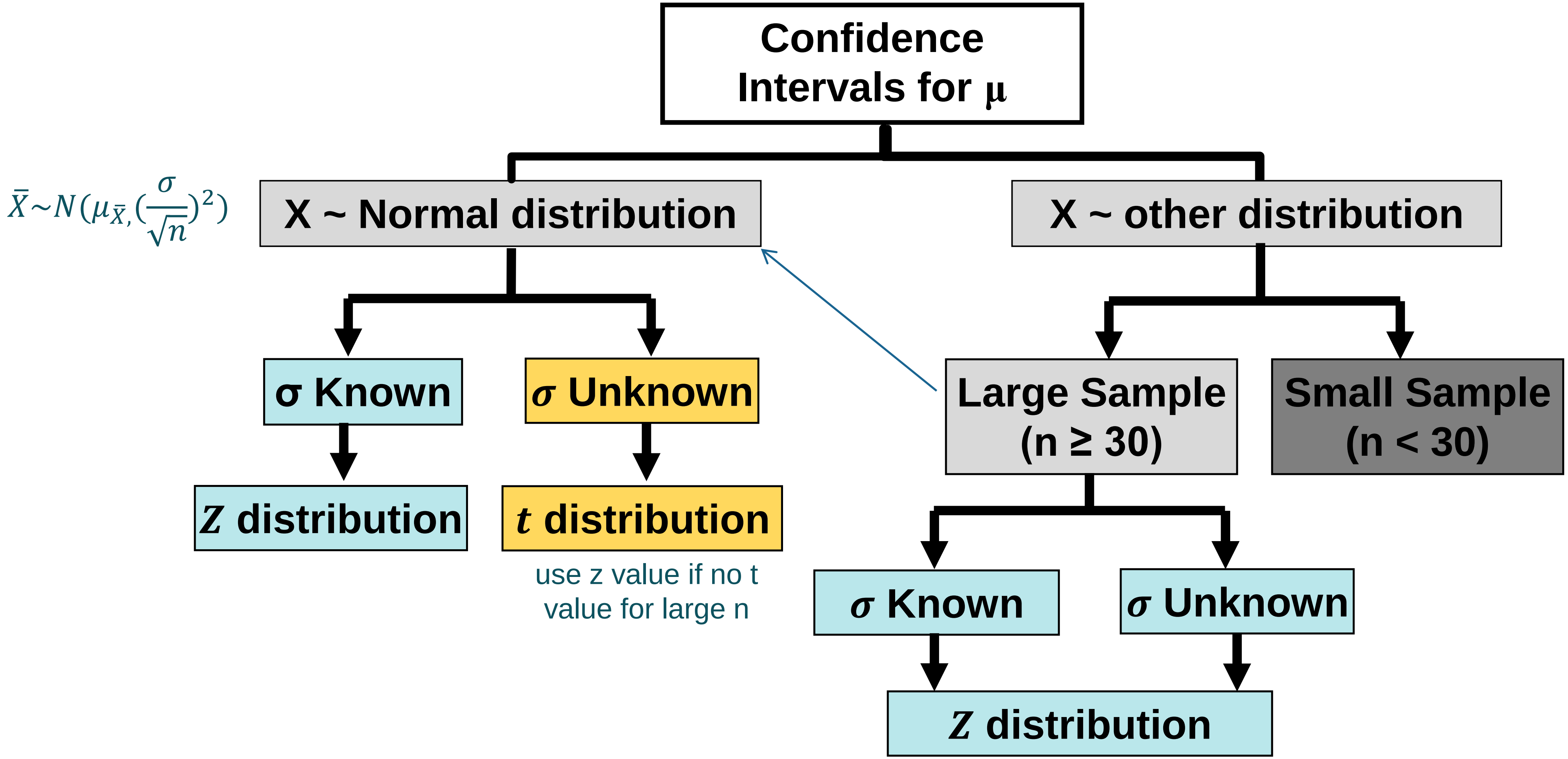
Key concept

T8 L5

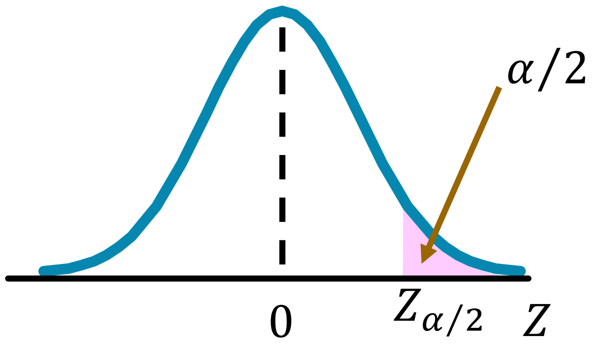
Point Estimates

	Population Parameters	Sample Statistics
Mean	μ	$\bar{X}$
Variance	σ^2	S^2
Proportion	p	$\hat{p}$

Confidence Interval for μ



Confidence Interval

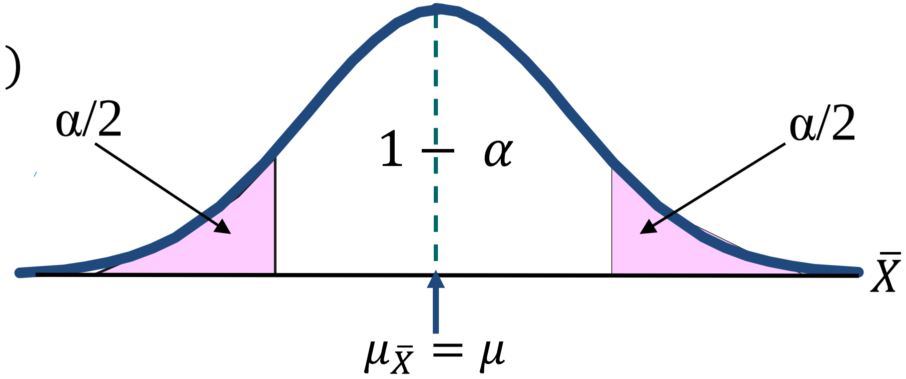


- Confidence interval
  - one sample
  - standard error = s.d. of sample
  - level of confidence $100(1-\alpha)\%$
- σ known : **Z distribution** $\boldsymbol{X \sim N(\mu, \sigma^2)}$
  - Z-value (Critical Value)

    $Z_{\alpha/2}$ is the value corresponding to an upper-tail probability of $\alpha/2$ from the standardized normal distribution
  - Sampling Error (Margin of Error)
    - $E = Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
- σ unknown : Student's **t distribution** $\boldsymbol{T \sim t(\nu)}$ S -> σ
  - Degrees of freedom in t-distribution
    - the number of values in the final calculation of a statistic that are free to vary
    - d.f. = total no. of observations - no. of parameters estimated as intermediate steps in the estimation of the parameter itself

Confidence Interval - Z

$\bar{X} \sim N(\mu_{\bar{X}}, (\frac{\sigma}{\sqrt{n}})^2)$



$$P\left(-Z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}\right) = 1-\alpha$$

$$\rightarrow P\left(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightarrow P\left(Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ge \mu-\bar{X} \ge -Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\rightarrow P\left(\boxed{\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}} \le \mu \le \boxed{\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha$$

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Standard Error, $\sigma_{\bar{X}}$

Sampling Error, E
(Also called margin of error)

Critical Value

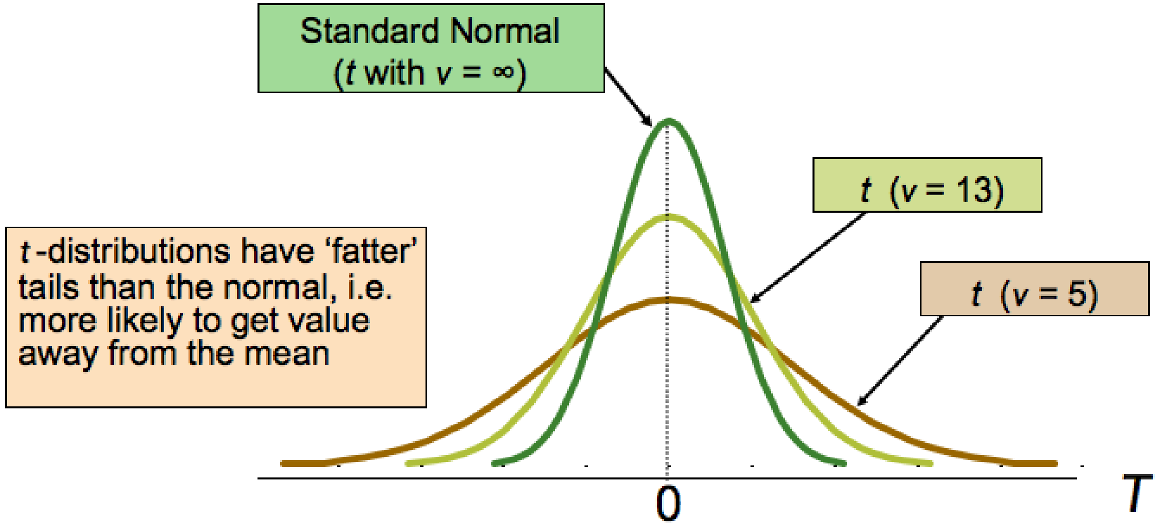
Confidence Interval - Z table - two-tail

Some common Z value and probabilities

Confidence level	P(Z>$Z_{\alpha/2}$)	$Z_{\alpha/2}$
80%	0.1	1.28
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575

Confidence Interval - t

- Mean & Standard Deviation
  - Mean = 0 for v >1, otherwise it is undefined
  - Standard deviation = $v/(v-2)$ for v > 2, = ∞ for $1<v\le 2$, otherwise undefined
- The shape of the density function
  - The theoretical range of T is infinite, i.e. $-\infty$ to $+\infty$
  - Bell shaped
  - Symmetric about T = 0
  - Median = mode = 0
  - As v increases, the density curve approaches the N(0, 1) curve



Confidence Interval - t

t distribution

$$\bar{X} \pm t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}$$

- with (n−1) degrees of freedom

Z distribution

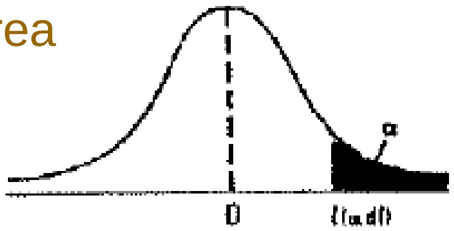
$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Confidence Interval - t table

The column gives the upper tail area

Critical Values of t

For a particular number of degrees of freedom, entry represents the critical value of t corresponding to a specified upper-tail area (α)



	Upper-Tail Areas					
Degrees of Freedom	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995

The row shows the degrees of freedom

The value within the table gives the t-value corresponding to a particular degrees of freedom and upper-tail area

At 7 degrees of freedom, $P(t > 2.3646) = 0.025$

Confidence Interval

- Interpretation
  - A relative frequency interpretation
    - In the long run, $100(1-\alpha)\%$ of all the confidence intervals that can be constructed will cover the unknown population parameter
  - A conventional interpretation
    - We are $100(1-\alpha)\%$ confident that the unknown population parameter lies between $\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $\bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

    - In other words, you got this interval by a method that gives correct results $100(1-\alpha)\%$ of the time

- A specific interval will either cover or not cover the population parameter
  - No probability involved in a specific interval -> cannot know whether the sample is one of the $100(1-\alpha)\%$ for which the interval catches μ, or one of the unlucky 5%

Confidence Interval

- Factors Affecting Interval Width (Precision)

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

data variation
sample size

$n \uparrow \rightarrow \sigma_{\bar{X}} \downarrow \rightarrow$ width of interval $\downarrow$

confidence level

$(1-\alpha) \uparrow \rightarrow$ $|Z$-value$|$ $\uparrow \rightarrow$ width of interval $\uparrow$

  - σ: unchanged
  - x: location, not width

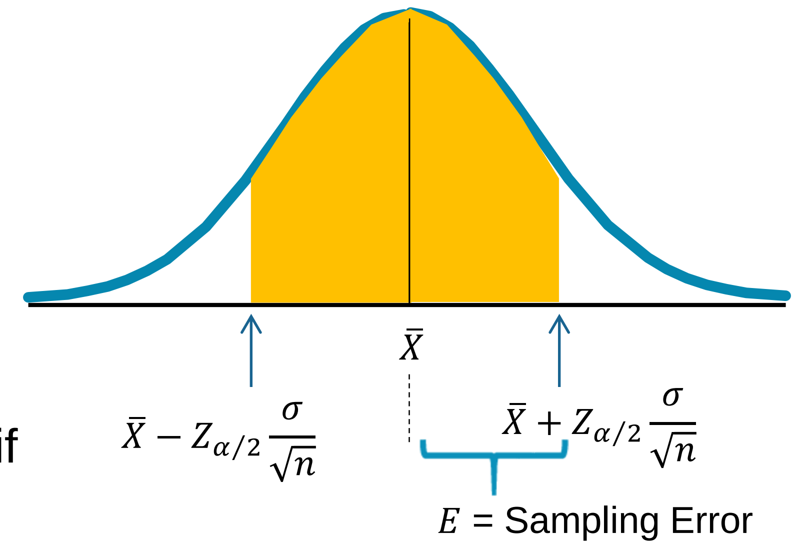
Determine Sample Size

Sampling error E: should be large enough to make the results credible

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

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

- Always round up: because sample size can only be integer; and if round down, the size will not be enough.



Standard deviation:

- when σ is unknown -> use S to replace σ
- when S is unknown -> use range/4 to estimate

Confidence Interval

P(z<a)	Z=a
0.01	-2.33
0.025	-1.96
0.05	-1.645
0.1	1.28
0.5	0
0.9	-1.28
0.95	1.645
0.975	1.96
0.99	2.33