Chapter 4

Recap Sheet

• What is a Confidence Interval

Problem: A point estimate gives one number for an unknown population value. Because of sampling randomness, $\bar{x}$ won't be exactly μ → doesn't say how close it is to μ.

⇒ **Idea**: A confidence interval gives a range of plausible values for the parameter, together with a confidence level (usually 90%, 95%, 99%)

Ex: "A 95% CI for μ is $[63.84, 65.08]$" means that if we repeated the whole sampling process many times and built a CI each time using the same method, about 95% of those integrals would contain the true μ.

What a CI tells us:
- A center (usually $\bar{x}$) and a half-width ("margin of error")
- Shorter intervals = more precise estimates

• Two-sided confidence interval for a mean when σ is known

We have a random sample $X_1, \dots, X_n$ for a normal population with unknown mean μ and known standard deviation σ.

1. Standardize the sample mean

The sample mean $\bar{X}$ has $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

So the standardized variable $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has a standard normal distribution $\mathcal{N}(0, 1)$.

2. Trap Z inside a central probability

For any $\alpha \in (0,1)$, let $z_{1-\frac{\alpha}{2}}$ be the upper $1-\frac{\alpha}{2}$ quantile of $\mathcal{N}(0,1)$ (e.g., $z_{0.025} = 1.96$ for 95%). Then

$$P(z_{\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

$z_{\alpha/2} < 0$, $z_{1-\frac{\alpha}{2}} > 0$ and $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$

3. Solve the double inequality for μ

Start with $\quad z_{\alpha/2} \leq \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\frac{\alpha}{2}}$

Multiply all parts by $\sigma/\sqrt{n}$: $\quad z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq z_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$

Subtract $\bar{X}$ and multiply by -1 (which flips the inequality):

$$\bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Result - Two-sided $100(1-\alpha)\%$ CI for μ: $\boxed{\bar{X} \pm z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}$

Ex : Given $n = 10$, known $\sigma = 1$ J, sample mean $\bar{x} = 64.46$ J, we build a 95 % CI ($\alpha = 0.05 \Rightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$)

Margin of error : $1.96 \cdot \frac{1}{\sqrt{10}} \approx 0.62$ J

CI : $64.46 \pm 0.62 \Rightarrow [63.84, 65.08]$ J

- Center : $\bar{x}$
- Margin of error (ME) : $z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
- Length : $2 z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ (total width of the interval)

What controls the length ?

1. Confidence level : higher confidence $\Rightarrow$ larger $z_{1-\frac{\alpha}{2}} \Rightarrow$ wider interval
2. Sample size n : Length $\propto \frac{1}{\sqrt{n}} \Rightarrow$ doubling n shrinks length by $\frac{1}{\sqrt{2}}$
3. Variability σ : larger $\sigma \Rightarrow$ wider interval

Planning sample size for a desired precision : If you want half-width (ME) at most h,

$$n = \left(\frac{z_{1-\frac{\alpha}{2}} \sigma}{h} \right)^2$$

Ex : Want a 95 % CI with ME $h = 0.30$ J and $\sigma = 1$:

$n = \left(\frac{1.96}{0.30} \right)^2 \approx 42.7 \Rightarrow 43$ specimens (round up)

- **One - sided confidence bounds for a mean when σ is known**

You only need a single bound on the mean, e.g.
- prove the mean is at least some level $\rightarrow$ lower bound
- prove the mean is at most some level $\rightarrow$ upper bound

From $Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$:

- Upper bound : want $P(\mu \leq U) = 1 - \alpha$

$$\mu \leq \bar{x} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \quad \text{(one - sided upper } 100(1-\alpha)\% \text{ bound)}$$

- Lower bound : want $P(L \leq \mu) = 1 - \alpha$

$$\bar{x} + z_{\alpha} \frac{\sigma}{\sqrt{n}} \leq \mu \quad \text{(one - sided lower } 100(1-\alpha)\% \text{ bound)}$$

⚠ One - sided uses z_α and not $z_{\alpha/2}$

- ## CI for a mean when σ is unknown (Student-t interval)

Data are a random sample from a normal population and the standard deviation σ is unknown → use the sample standard deviation $S^* = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)}$

- t - statistic : $T = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ (if population is Normal)

Two - sided $100(1-\alpha)\%$ CI for μ : Let $t_{1-\frac{\alpha}{2}, n-1}$ be the upper $1-\frac{\alpha}{2}$ quantile distribution with $n-1$ degrees of freedom (df).

$$\bar{x} \pm t_{1-\frac{\alpha}{2}, n-1} \frac{S^*}{\sqrt{n}}$$

One - sided bounds :

- Upper $100(1-\alpha)\%$ bound : $\mu \leq \bar{x} + t_{1-\alpha, n-1} \frac{S^*}{\sqrt{n}}$
- Lower $100(1-\alpha)\%$ bound : $\bar{x} + t_{\alpha, n-1} \frac{S^*}{\sqrt{n}} \leq \mu$

Ex : Suppose $n = 10$, sample mean $\bar{x} = 64.46$ J and sample sd $s = 1.00$ J

For a 95 % CI, $t_{0.975,9} = 2.262$.

Margin $= 2.262 \times \frac{1.00}{\sqrt{10}} \approx 0.716$ J

95 % CI : $64.46 \pm 0.716 = [63.74, 65.18]$ J

- ## Confidence interval for a variance (and std) of a Normal population

You want a CI for the population variance σ^2 (or σ) and data are a random sample from a Normal population.

With $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ and $S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, we have :

$$\frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2_{n-1}$$

Two - sided $100(1-\alpha)\%$ CI for σ^2 :

$$\left[\frac{(n-1)S^{*2}}{\chi^2_{1-\alpha/2, n-1}} , \frac{(n-1)S^{*2}}{\chi^2_{\alpha/2, n-1}} \right]$$

One - sided bounds :

- Upper : $\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\alpha, n-1}}$
- Lower : $\frac{(n-1)S^2}{\chi^2_{1-\alpha, n-1}}$

Ex : Suppose $n = 10$ specimens, sample std $s = 1.00$ (so $s^2 = 1.00$).
For a 95 % CI, df = 9.
Common chi - square quantiles : $\chi^2_{0.025,9} \approx 2.700$, $\chi^2_{0.975,9} \approx 19.023$

Two-sided 95% CI for σ^2 : $\left[\frac{9\times 1.00}{19.023}, \frac{9\times 1.00}{2.700}\right] = [0.473, 3.333]$

CI for σ : $[\sqrt{0.473}, \sqrt{3.333}] = [0.688, 1.826]$

- **Confidence interval for a population proportion p**

We run n independant, identical trials (each success / fail) with true success probability p.
Let X = number of successes $\sim B(n,p)$
The sample proportion (point estimate) is $\hat{p} = \frac{X}{n}$

When the usual CI works (large-sample rule) : Use the normal approximation when both counts are reasonably large :

$$n\hat{p} \gtrsim 10 \quad \text{and} \quad n(1-\hat{p}) \gtrsim 10$$

Standard two-sided $100(1-\alpha)\%$ CI :

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

One-sided bounds :

- Upper bound : $p \leq \hat{p} + z_{1-\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Lower bound : $p \geq \hat{p} + z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Ex : You observe $X = 230$ successes out of $n = 400$
- $\hat{p} = 230/400 = 0.575$
- Check large-sample rule : $n\hat{p} = 230 \geq 10$, $n(1-\hat{p}) = 170 \geq 10$

95% CI ($z_{0.975} = 1.96$) :

$SE = \sqrt{\frac{0.575 \times 0.425}{400}} \approx 0.02472$

$ME = 1.96 \times 0.02472 \approx 0.04845$

$CI = 0.575 \pm 0.04845 \Rightarrow [0.527, 0.623]$

- **Roadmap to construct a CI**

1. Parameter I want to estimate
2. Estimator of the parameter
3. Law of the estimator
4. Probability (e.g. $P(-z_{\alpha/2} \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1-\alpha$
5. Interval formula
6. Calculation

• IC à 95 % pour μ, σ^2 connue

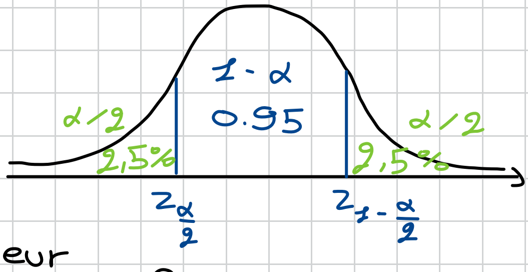
$\mu \in [$

$P(\mu \in I) = 95\% = 1 - \alpha$ donc $\alpha = 5\% = 0{,}05$

$\dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1)$

L'intervalle tq : $P\left(z_{\frac{\alpha}{2}} \leq \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \leq z_{1-\frac{\alpha}{2}}\right) = 0\ 95$

$\Rightarrow \mu \in [\bar{x} + z_{\frac{\alpha}{2}} \,;\, \bar{x} + z_{1-\frac{\alpha}{2}}]$



Pour trouver les valeurs de $z_{\frac{\alpha}{2}}$ et $z_{1-\frac{\alpha}{2}}$:

- Table 1 : on cherche à l'intérieur la valeur $1 - \frac{\alpha}{2}$ et sur les bords on a le $z_{1-\frac{\alpha}{2}}$ associé

  et on déduit $z_{\frac{\alpha}{2}} = - z_{1-\frac{\alpha}{2}}$