10: Hypothesis Testing and Confidence Intervals

Naijia Liu

Spring 2025

Interval estimation - what and why?

- $\hat{\tau} = \overline{Y}_n - \overline{X}_n$ is our best guess about $\tau = \mu_y - \mu_x$
- But $\mathbb{P}(\hat{\tau} = \tau) = 0$!
- Alternative: produce a range of plausible values instead of one number.
  - Hopefully will increase the chance that we've captured the truth.
- We can use the distribution of estimators to derive these intervals.

Definitions

- **Interval estimator** of θ is an interval between two statistics $C = [L, U]$.
  - $L = L(X_1, \ldots, X_n)$ and $U = U(X_1, \ldots, X_n)$ are functions of the data.
  - An estimator just like $\overline{X}_n$ but with two values.
  - Goal: to infer that C covers or contains the true value.
- **Coverage probability** of $C = [L, U]$ is the probability that C covers the true value θ.
- In math, $\mathbb{P}(L \leq \theta \leq U) = \mathbb{P}(\theta \in C)$
- Important: interval is the random quantity, not the parameter.

What is a confidence interval?

Definition

A $1 - \alpha$ **confidence interval** for a population parameter θ is an interval estimator $C = (L, U)$ with coverage probability $1 - \alpha$.

- The random interval (L, U) will contain the truth $1 - \alpha$ of the time.
- Ideally, we'd want a 100% confidence interval, but usually not possible.
- Extremely useful way to represent our uncertainty about our estimate.
  - Shows a range of plausible values given the data.

Simple confidence intervals

- Estimator $\hat{\theta}$ for θ with estimated standard error $\widehat{\text{se}}[\hat{\theta}]$.
- Quick-and-dirty 95% confidence interval:

$$C = \left[\hat{\theta} - 2\widehat{\text{se}}(\hat{\theta}),\ \hat{\theta} + 2\widehat{\text{se}}(\hat{\theta})\right]$$

- More formal, normal-based $1 - \alpha$ confidence interval:

$$C = \left[\hat{\theta} - z_{1-\alpha/2}\widehat{\text{se}}(\hat{\theta}),\ \hat{\theta} + z_{1-\alpha/2}\widehat{\text{se}}(\hat{\theta})\right]$$

  - $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal.
- Student-t-based $1 - \alpha$ confidence interval:

$$C = \left[\hat{\theta} - q_{1-\alpha/2}\widehat{\text{se}}(\hat{\theta}),\ \hat{\theta} + q_{1-\alpha/2}\widehat{\text{se}}(\hat{\theta})\right]$$

  - $q_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the student t with df $= r$.

Deriving confidence intervals

- How do we know the coverage of these confidence intervals?
- Sample mean: $\overline{X}_n$ with X_i i.i.d. $\mathcal{N}(\mu, \sigma^2)$ with $\widehat{\text{se}} = s/\sqrt{n}$.

$$T = \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

- T is a **pivotal quantity**: distribution doesn't depend on θ.
- If $q_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the t_{n-1}, we have

$$\mathbb{P}\left(-q_{1-\alpha/2} \leq \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \leq q_{1-\alpha/2}\right) = 1 - \alpha$$

Finding the critical values

- How do we figure out what $q_{1-\alpha/2}$ will be?
- Intuitively, we want the q values that puts $\alpha/2$ in each of the tails.
- $$\mathbb{P}(q_{\alpha/2} \leq T \leq q_{1-\alpha/2}) = 1 - \alpha$$
  - Because t is symmetric, we have $q_{\alpha/2} = -q_{1-\alpha/2}$
  - If $G(t)$ is the c.d.f. of T, then we have $q_{1-\alpha/2} = G^{-1}(1 - \alpha/2)$

Deriving the interval

- Let's work backwards to derive the confidence interval:

$$
\begin{aligned}
1-\alpha &= \mathbb{P}\left(-q_{1-\alpha/2} \leq \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \leq q_{1-\alpha/2}\right) \\
&= \mathbb{P}\left(-q_{1-\alpha/2} \times \frac{s}{\sqrt{n}} \leq \overline{X}_n - \mu \leq q_{1-\alpha/2} \times \frac{s}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(-\overline{X}_n - q_{1-\alpha/2} \times \frac{s}{\sqrt{n}} \leq -\mu \leq -\overline{X}_n + q_{1-\alpha/2} \times \frac{s}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\overline{X}_n - q_{1-\alpha/2} \times \frac{s}{\sqrt{n}} \leq \mu \leq \overline{X}_n + q_{1-\alpha/2} \times \frac{s}{\sqrt{n}}\right)
\end{aligned}
$$

- Lower bound: $\overline{X}_n - q_{1-\alpha/2} s/\sqrt{n}$
- Upper bound: $\overline{X}_n + q_{1-\alpha/2} s/\sqrt{n}$
  - For 95% confidence interval with $n = 100$, $q_{1-\alpha/2} = 1.98$
- Bounds are random! Not μ!

Asymptotic confidence intervals

- What about the $1-\alpha$ normal confidence interval:

$$C = \left[\hat{\theta} - z_{1-\alpha/2}\widehat{\mathsf{se}}(\hat{\theta}),\ \hat{\theta} + z_{1-\alpha/2}\widehat{\mathsf{se}}(\hat{\theta})\right]$$

- Asymptotically valid if our estimator is asymptotically normal so that:

$$\frac{\hat{\theta}_n - \theta}{\widehat{\mathsf{se}}(\hat{\theta})} \xrightarrow{d} \mathcal{N}(0,1)$$

- Then as $n \to \infty$

$$\mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\hat{\theta}_n - \theta}{\widehat{\mathsf{se}}(\hat{\theta})} \leq z_{1-\alpha/2}\right) = \mathbb{P}(\theta \in C) \to 1-\alpha$$

- Again, $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ (`qnorm` in R)

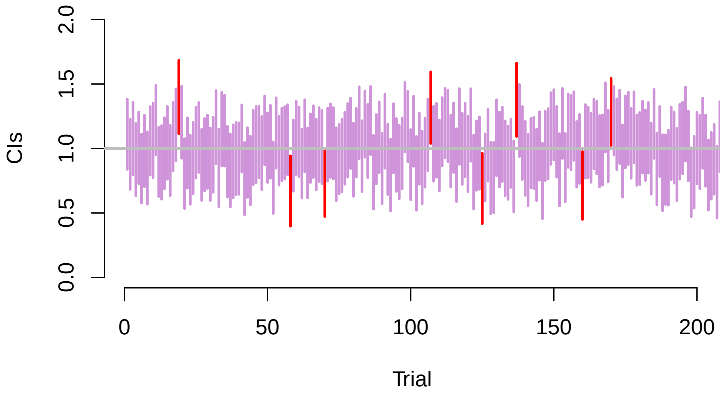
Interpreting the confidence interval

- **Caution:** a common **incorrect** interpretation of a confidence interval:
  - "I calculated a 95% confidence interval of [0.05, 0.13], which means that there is a 95% chance that the true difference in means is in that interval."
  - This is **WRONG**.
- The true value of the population mean, μ, is **fixed**.
  - It is either in the interval or it isn't—there's no room for probability at all.
- The randomness is in the interval: $\overline{X}_n \pm 1.96 \times s/\sqrt{n}$
- Correct interpretation: **across 95% of random samples, the constructed confidence interval will contain the true value.**

Simulations

```
######################################
Fold code 10000
cover <- rep(0, times = sims)
low.bound <- up.bound <- rep(NA, times = sims)
for(i in 1:sims){
  draws <- rnorm(500, mean = 1, sd = sqrt(10))
  low.bound[i] <- mean(draws) - sd(draws) / sqrt(500) * 1.96
  up.bound[i] <- mean(draws) + sd(draws) / sqrt(500) * 1.96
  if (low.bound[i] < 1 & up.bound[i] > 1) {
    cover[i] <- 1
  }
}
mean(cover)
```

Simulations



Inverting a hypothesis test

- 95% confidence interval: $\overline{X}_n \pm 1.96 \times s/\sqrt{n}$
- **CI/Test duality:** A $1-\alpha$ confidence interval contains all null hypotheses that we would not reject with a α-level test.
- Test of the null $H_0 : \mu = \mu_0$ at size α and reject when $|T| > z_{1-\alpha/2}$ where

$$T = \frac{\overline{X}_n - \mu_0}{s/\sqrt{n}}$$

- Reject when $\mu_0 > \overline{X}_n + z_{1-\alpha/2}\, s/\sqrt{n}$ or $\mu_0 < \overline{X}_n - z_{1-\alpha/2}\, s/\sqrt{n}$
    - $\rightsquigarrow$ reject any null outside the 95% confidence interval at size $\alpha = 0.05$
- CIs are a range of plausible values in the sense we cannot reject them as null hypotheses.