Hypotheses and Tests

Tsering Sherpa

2025

Confidence Intervals

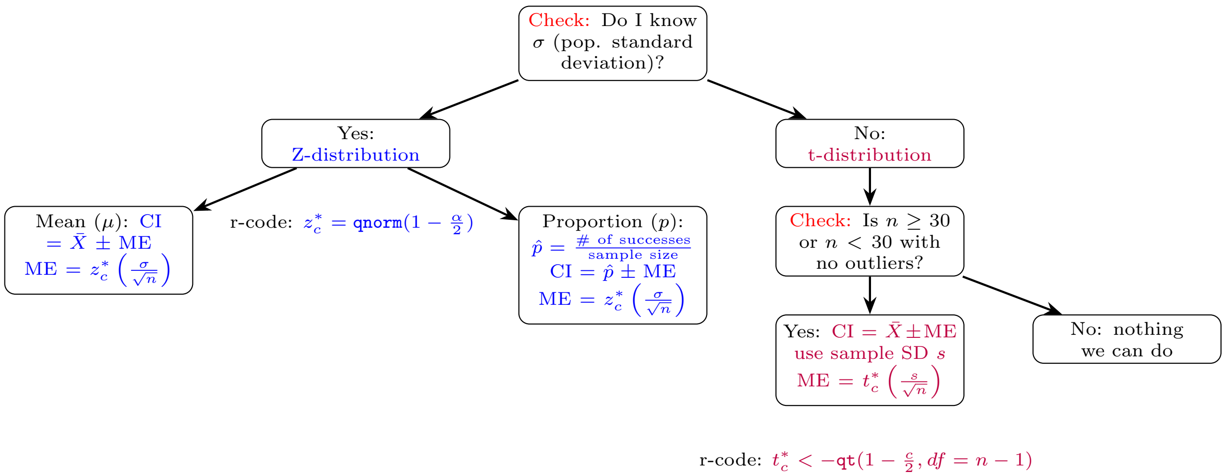
A **confidence interval (CI)** is a range of values used to estimate a population parameter. A $c\%$ level of confidence means that if we repeated the estimation procedure many times, about $c\%$ of the resulting intervals would contain the true population parameter. Either the parameter is in the confidence interval or not, so we do NOT use the word "probability", we can say we are confident.

- **Incorrect**: "There is a 95% chance that μ is in the interval (22.24, 23.56)."
- **Correct**: "If we were to repeat the sampling process many times, approximately 95% of the constructed intervals would contain the true mean μ."
- Acceptable: "We estimate with 95% confidence that the true population mean lies within (22.24, 23.56)."

Conditions for Confidence Intervals

- If $n < 30$: Data should come from a nearly normal distribution with no outliers.
- If $n \geq 30$: Central Limit Theorem justifies approximate normality of $\bar{X}$, even if original data isn't normal.

Which Distribution to Use



Critical Values

- The critical value z_c or $t_{c,\text{df}}$ corresponds to the desired confidence level.
- Higher confidence levels yield larger critical values.

Hypotheses and Test Direction

The **null hypothesis** H_0 reflects no effect or no difference. The **alternative hypothesis** H_a reflects the effect or direction you are testing for.

Writing the Hypotheses

- H_0: μ has no effect/difference (the claim about the population)
- H_a: μ does reflect an effect/direction (you test this alternative).

Examples

Understanding hypothesis testing helps us evaluate data, think critically, and remain skeptical rather than jumping to conclusions. It provides a systematic framework for making decisions under uncertainty. Here are some real-world examples:

1. **Medical Research:** Does a new drug reduce symptoms compared to a placebo?
   Null hypothesis: the average effect of the drug equals that of the placebo, $H_0 : \mu_{\text{drug}} = \mu_{\text{placebo}}$.
   We test whether the observed difference in outcomes is statistically significant to infer if the drug actually works.

2. **Sports Analytics:** Does a new basketball technique improve free-throw accuracy?
   Null hypothesis: $H_0 : p_{\text{new}} = p_{\text{old}}$,
   Alternative hypothesis: $H_a : p_{\text{new}} > p_{\text{old}}$.
   We compare the proportions of successful free throws before and after implementing the technique.

3. **Classroom Examples:** Use examples from class to illustrate cases where students test whether observed outcomes differ significantly from expectations (e.g., exam score comparisons, A/B testing with click rates, etc.).

Table 1: Types of Hypothesis Tests

Situation	H_0	H_a	**Direction**	**Meaning**
Mean equals a value	$\mu = \mu_0$	$\mu \neq \mu_0$	Two-tailed	Difference in either direction
Mean greater than a value	$\mu \leq \mu_0$	$\mu > \mu_0$	Right-tailed	Directional hypothesis
Mean less than a value	$\mu \geq \mu_0$	$\mu < \mu_0$	Left-tailed	Directional hypothesis
Proportions equal	$p_1 = p_2$	$p_1 \neq p_2$	Two-tailed or One-tailed	Can be directional or not

Use a one-tailed test if you have a directional hypothesis, e.g., $p_1 > p_2$ or $p_1 < p_2$. Otherwise, use a two-tailed test for a general difference.

(Standardized) Test Statistics and Critical Values

The terms **test statistic** and **standardized test statistic** are related but distinct in hypothesis testing. A **test statistic** is a quantity computed from sample data that summarizes the evidence against the null hypothesis. Below is a graph to help you visualize this.

A **standardized test statistic**, such as the z-score, is calculated to measure how many standard deviations the sample mean $\bar{X}$ is from the population mean μ, assuming the null hypothesis is true:

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

This formula shows how far the sample mean $\bar{X}$ is from the population mean μ (we claim under the null) in standard deviations. This standardization allows for comparison across different tests by placing results on a common reference distribution, such as the standard normal distribution.

Decision Rule

If the computed test statistic satisfies $|z| > z_{\alpha/2}$, we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

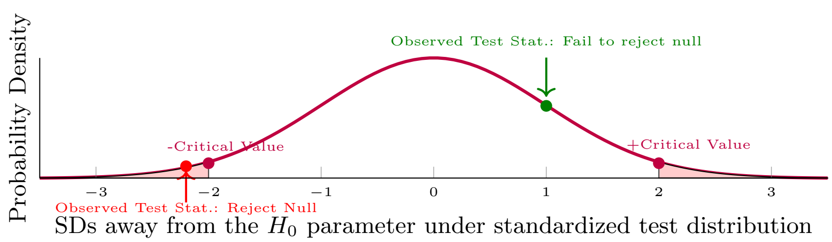
To determine whether the null hypothesis should be rejected, the standardized test statistic is compared to a **critical value** determined by the chosen significance level α. For example, at the 5% significance level in a two-tailed test, the critical values are:

$$z_{\alpha/2} = \pm 1.96$$

Example: Critical Values and Rejection Regions (for $\alpha = 0.05$)

Test Type	**Rejection Region**	**Critical Value**
Two-tailed	$\lvert z\rvert > z_{\alpha/2}$	± 1.96
Left-tailed	$z < -z_\alpha$	$z < -1.645$
Right-tailed	$z > z_\alpha$	$z > 1.645$

Figure 1: Standardized Test Statistic Under Null Hypothesis Distribution H_0



Area: The pink area represents α. Think of this when you are applying one vs two-sided tests.

You should be able to tell why we need to standardize the observed test statistic to use our tests. Look at the x-axis and y-axis and see what happens if you just try to plot a test statistic before standardizing it.

Table 2: Types of Z and T-Tests

Test Type	**Test**	**One-Sample**	**Two-Sample**	**Paired**
Z-Test	Standardized Mean Statistic	$z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$z = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	–
T-Test	Standardized Mean Statistic	$t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$	$t = \frac{\bar{X}_1-\bar{X}_2}{\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}}$	$t = \frac{\bar{d}}{s_d/\sqrt{n}}$
Z-Test (Proportions)	Standardized Proportion Statistic	$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$	–

In the paired t-test: $\bar{d}$ is the mean of the differences between paired observations, s_d is the standard deviation of those differences, and n is the number of pairs.