Testing Statistical Hypotheses

STAT305 - Fall 2025

Lecture 01: Foundations (H_0/H_1, p-values, power)

Dr. Zaid T. Al-Khaledi

Department of Statistics & Informatics
College of Computer Science & Mathematics
University of Mosul

October 1, 2025

Learning outcomes (today)

By the end of this lecture, students will be able to:

1. Tell the difference between **parameters** (population) and **statistics** (sample). Explain **standard error (SE)** and sampling distributions.
2. Write clear $\mathbf{H}_0$ and $\mathbf{H}_1$ (one-sided or two-sided) for a question.
3. Define and interpret **p-value**, **confidence interval (CI)**, **Type I/II errors**, and **power**.
4. Follow the testing steps: assumptions → statistic/SE → p-value/CI → decision → interpretation.
5. Report results with an estimate, a CI, a p-value, and a simple **effect size**.

Parameters vs Statistics (what and why)

- **Parameter** (μ, σ^2, p): a fixed but unknown number that describes the population.
- **Statistic** ($\bar{X}$, S^2, $\hat{p}$): a number from the sample; it changes from sample to sample (it is *random*).
- **Estimator**: a statistic used to estimate a parameter (e.g., $\bar{X}$ estimates μ).
- **Sampling distribution**: the distribution of a statistic across many repeated samples.
- **Standard error (SE)**: the standard deviation of the sampling distribution. It measures the precision of the estimate.

We do not see the parameter. We only see data, and we compute statistics. The *sampling distribution* tells us: (1) how much the statistic moves from sample to sample (this gives the *SE* and the width of a *CI*); and (2) what values are typical if H_0 is true (so we can judge whether our observed statistic is rare and compute a *p-value*).

Sampling distributions & SE (quick facts)

- For large samples, the sample mean is approximately normal: $\bar{X} \approx \mathcal{N}(\mu, \sigma^2/n)$ (Central Limit Theorem).
- If σ is known: $SE(\bar{X}) = \sigma/\sqrt{n}$. If σ is unknown: use $S/\sqrt{n}$ and t-methods.
- For a proportion: $\hat{p} = \frac{1}{n}\sum I(X_i = 1)$, and $SE(\hat{p}) \approx \sqrt{\hat{p}(1-\hat{p})/n}$.

Bigger samples give smaller SE, but not twice as good. SE goes like $1/\sqrt{n}$. If you double n, SE becomes $1/\sqrt{2} \approx 0.71$ (about 29% smaller, not half). With very large n, even tiny effects can be "statistically significant". So always report an *effect size* and a *CI* to show size and precision.

Null and Alternative Hypotheses

Key ideas

- **Null hypothesis ($\mathbf{H}_0$)**: "no change / no effect" or a specific value (e.g., $\mu = \mu_0$). This is the claim we test.
- **Alternative hypothesis ($\mathbf{H}_1$)**: there *is* a change or effect (e.g., $\mu \neq \mu_0$, $\mu > \mu_0$, or $\mu < \mu_0$).
- **Simple vs composite**: $H_0 : \mu = \mu_0$ is *simple* (one value). $H_1 : \mu \neq \mu_0$ is *composite* (many values).

Choose H_0 and H_1 from the research question, *before* you look at the data. H_0 tells us what values of the test statistic are typical when there is no effect. We then check if our observed statistic is far from these typical values. If yes, we have evidence against H_0.

One-sided vs Two-sided tests (how to choose)

Question	Hypotheses
"Has the mean changed?"	$H_0 : \mu = \mu_0; \quad H_1 : \mu \neq \mu_0$ (two-sided)
"Is the mean higher?"	$H_0 : \mu \leq \mu_0; \quad H_1 : \mu > \mu_0$ (right-tailed)
"Is the mean lower?"	$H_0 : \mu \geq \mu_0; \quad H_1 : \mu < \mu_0$ (left-tailed)

Use **two-sided** by default. Use a **one-sided** test only if the other direction does not matter or is impossible, and decide this *before* analysis. One-sided tests have more power in the chosen direction, but they cannot detect effects in the other direction.

Test statistic, p-value, and CI

Test statistic & p-value

- Typical form:
$$T = \frac{\text{estimate} - \text{null value}}{\text{SE}}.$$
- **p-value**: the probability, *assuming* H_0, to see a result as extreme as (or more than) ours.
- Decision rule: reject H_0 if p-value $\leq \alpha$ (e.g., 0.05).

Confidence interval (CI)

- A range of reasonable values for the parameter.
- Width shows precision (depends on SE, n, and variability).
- **Link**: a two-sided test at level α rejects $H_0 : \theta = \theta_0$ exactly when the $(1 - \alpha)$ CI does *not* contain θ_0.

Good practice: Report the estimate with its CI and the p-value together.

Type I and Type II Errors (at a glance)

	H_0 **true**	H_0 **false** (H_1 **true**)
Rejected	**Type I error** False positive Probability = α	**Correct decision** True positive Probability = $1 - \beta$
Not rejected	**Correct decision** True negative Probability = $1 - \alpha$	**Type II error** False negative Probability = β

Power = $1 - \beta$ (chance to reject H_0 when H_1 is true).

Testing workflow (simple recipe)

1. **State** H_0/H_1 and choose one- or two-sided; fix α.
2. **Check assumptions** (design/independence; normality or large n; equal variance if required).
3. **Compute** the estimate, SE, and the test statistic T (e.g., z or t).
4. **Decision (two equivalent ways):**
   - **Critical/tabulated value method:**
     - Two-sided: reject H_0 if $|T| > t_{1-\alpha/2,\ \mathrm{df}}$ (or $|z| > z_{1-\alpha/2}$).
     - Right-tailed: reject if $T > t_{1-\alpha,\ \mathrm{df}}$; Left-tailed: reject if $T < -t_{1-\alpha,\ \mathrm{df}}$.
   - **p-value method:** reject H_0 if $p \le \alpha$; otherwise **fail to reject**.
5. **Also report** a **CI** and an **effect size**; give a plain-language conclusion.

Worked example: one mean (two-sided)

Goal. Check if the mean fill differs from 500 ml. From past runs the process SD is known: $\sigma = 12$ ml. We take a sample of $n = 36$ bottles with mean $\bar{x} = 504$ ml. Use $\alpha = 0.05$ (two-sided).

Step 1: $H_0 : \mu = 500;\quad H_1 : \mu \neq 500.$ **Step 2:** $SE = \sigma/\sqrt{n} = 12/6 = 2.$

Step 3: $z = \dfrac{\bar{x} - \mu_0}{SE} = \dfrac{504 - 500}{2} = 2.00.$

Step 4a (Critical/tabulated). For $\alpha = 0.05$ two-sided, $z_{1-\alpha/2} = z_{0.975} = 1.96$. Since $|z| = 2.00 > 1.96$, **reject** H_0.

Step 4b (p-value). Two-sided $p = 2\,[1 - \Phi(2.00)] \approx 0.0455$. Since $p = 0.0455 \leq 0.05$, **reject** H_0. (Both methods agree.)

Step 5: CI & effect size. 95% CI = $\bar{x} \pm 1.96 \times SE = 504 \pm 3.92 = (500.1,\ 507.9)$ (does not contain 500).
Cohen's $d = (504 - 500)/12 = 0.33$ (small-moderate).

The average fill is slightly *above* 500 ml; the difference is statistically significant but small.

If the sample were weaker (fail to reject)

Keep $\sigma = 12$, $n = 36$, $\alpha = 0.05$ two-sided, but suppose $\bar{x} = 503$ ml.

$SE = 2, \quad z = (503 - 500)/2 = 1.50.$
Critical/tabulated: $|z| = 1.50 < 1.96 \Rightarrow$ **fail to reject** H_0.
p-value: $p = 2\,[1 - \Phi(1.50)] \approx 0.133 > 0.05 \Rightarrow$ **fail to reject** H_0.
95% CI: $503 \pm 1.96 \times 2 = (499.1,\ 506.9)$ (contains 500).

With this sample, we do not have enough evidence to say the mean differs from 500 ml. (Effect might be small, or the study may need a larger sample.)

Common pitfalls & quick checklist

Checklist

1. Clear H_0/H_1 and tail direction, set before analysis (α fixed).
2. Assumptions checked (design, normality/large n, variances if needed).
3. Show the test statistic and *either* decision method (critical value or p-value). Both should agree.
4. Add a CI and an effect size; conclude in plain language linked to the question.