Lecture Notes: Hypothesis Testing & Statistical Inference

PSTAT 5A

July 30, 2025

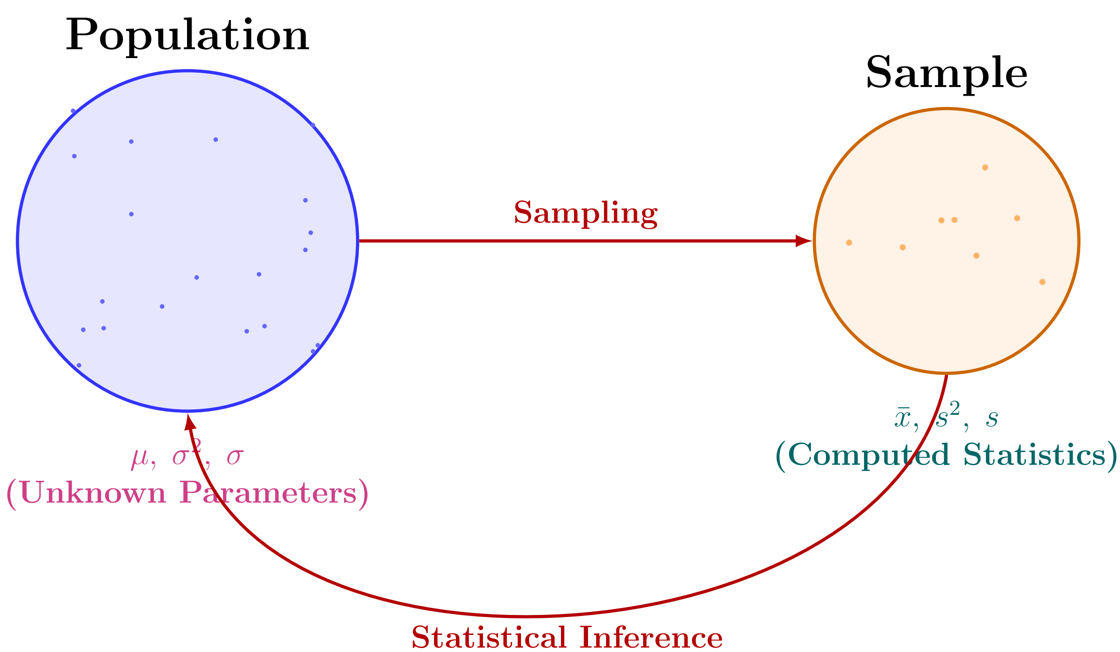
Contents

1 From Population to Inference

In statistics, we start with an entire **population** of size N and draw a smaller **sample** of size n at random. Our objective is to use what we *observe* in the sample to learn about what we *cannot observe* in the population.

- **Population** (N): every UCSB student's height, GPA, etc.
- **Sample** (n): a subset chosen independently and at random.
- **Parameters**: unknown numbers that characterise the population, e.g. μ (mean), σ^2 (variance), σ (standard deviation).
- **Statistics**: computable numbers from the sample, e.g. $\bar{x}$, s^2, s, used as *estimators*.

Goal: Use statistics to make reliable statements about the parameters.



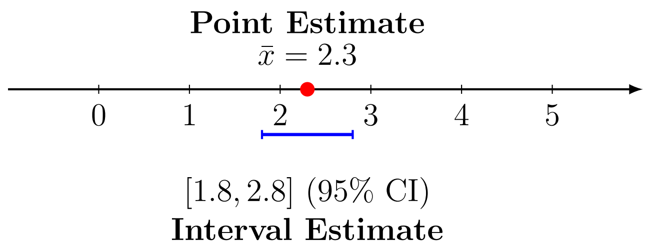
The workflow involves:

1. **Sampling**: Move from the population to a sample.
2. **Inference**: Use statistics to estimate parameters and quantify uncertainty.

2 Point Estimates vs. Interval Estimates

Definition 2.1 (Types of Statistical Estimates)**.** *Statistical procedures for unknown parameters fall into two categories:*

- ***Point estimates****: Return a* single *numerical value (one point on the number line)*
- ***Interval estimates****: Return a* range *of plausible values (confidence intervals)*



Parameter	Point Estimator	Example Output
Mean μ	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	72.4 cm
Variance σ^2	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	81.0 cm^2
Proportion p	$\hat{p} = \frac{k}{n}$ (where k = successes)	0.37

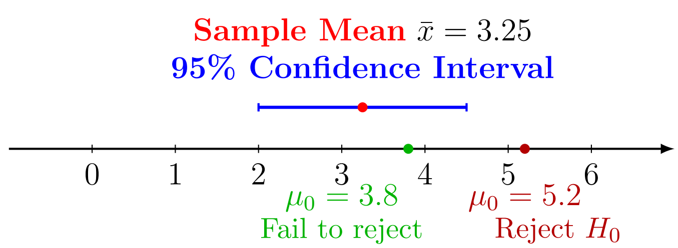
Table 1: Common point estimators and their typical outputs.

3 The Bridge: Confidence Intervals $\leftrightarrow$ Hypothesis Tests

Understanding the relationship between confidence intervals and hypothesis tests is crucial:

Key Relationships

- **Point Estimate** $\bar{x}$: Single best guess of μ
- **Confidence Interval**: Range of *plausible* values for μ
- **Hypothesis Test**: Asks if *one specific* value μ_0 is plausible at significance level α



4 Introduction to Hypothesis Testing

Hypothesis testing asks whether the data contradict a *default claim* about a population parameter. Rejecting that claim requires sufficiently strong sample evidence.

4.1 Core Vocabulary

Hypothesis Testing Terminology

H_0 **(Null)** "No effect" or status-quo value we assume true until proven otherwise

H_a **(Alternative)** What we hope to support; specifies direction ($>$, $<$) or simply $\neq$

Test Statistic Single number (e.g., z, t, χ^2) quantifying distance between sample estimate and H_0

p**-value** Probability, *if* H_0 *were true*, of obtaining a test statistic at least this extreme

α **(Significance)** Pre-chosen Type I error rate (commonly 0.05 or 0.01)

Decision Reject H_0 if $p \leq \alpha$ (or test statistic falls in critical region)

5 Understanding P-values

The p-value is one of the most important concepts in hypothesis testing, yet it's often misunderstood. Let's break it down clearly.

What is a P-value?

The p-value is the probability of observing data at least as favorable to the alternative hypothesis as our current data set, **if the null hypothesis were true**.
In simpler terms: "If H_0 is actually true, what's the chance of getting results this extreme or more extreme?"

5.1 P-value Properties and Interpretation

Key Properties of P-values

- **Range:** $0 \leq p \leq 1$ (it's a probability)
- **Small p-value:** Strong evidence against H_0
- **Large p-value:** Weak evidence against H_0
- **Decision rule:** Reject H_0 if $p \leq \alpha$

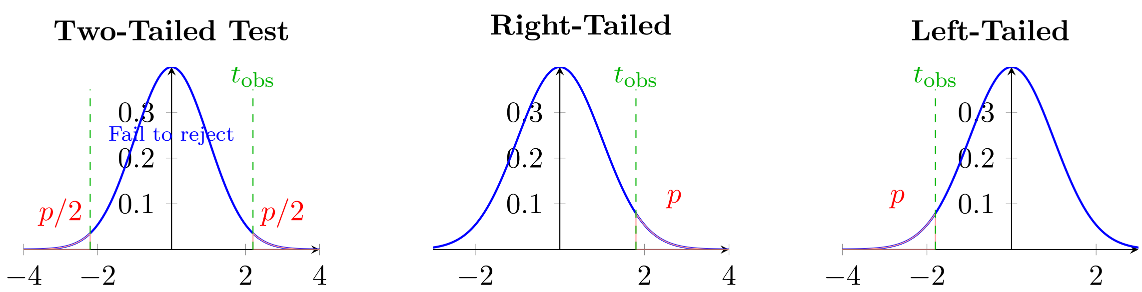
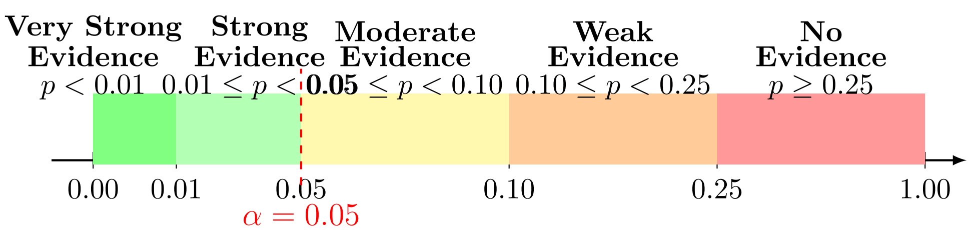


Figure 1: Visual representation of p-values for two- and one-tailed tests.

5.2 P-value Strength Guide



5.3 Common P-value Misconceptions

What P-values Are NOT!

WRONG Interpretations:

- ✗ "The probability that H_0 is true"
- ✗ "The probability that H_a is true"
- ✗ "The probability of making an error"
- ✗ "The probability the results are due to chance"

CORRECT Interpretation:

- ✓ "If H_0 were true, the probability of observing data this extreme or more extreme"

5.4 P-value Examples with Context

Example 5.1 (Interpreting Different P-values)**.** *Consider testing whether a new teaching method improves test scores:*

Scenario 1: $p = 0.003$

- ***Meaning:*** *If the new method had no effect, there's only a 0.3% chance of seeing improvement this large or larger*
- ***Conclusion:*** *Very strong evidence that the method works*
- ***Decision:*** *Reject H_0 (method has no effect)*

Scenario 2: $p = 0.08$

- ***Meaning:*** *If the new method had no effect, there's an 8% chance of seeing improvement this large or larger*
- ***Conclusion:*** *Moderate evidence, but not quite significant at $\alpha = 0.05$*
- ***Decision:*** *Fail to reject H_0 (borderline case)*

Scenario 3: $p = 0.35$

- ***Meaning:*** *If the new method had no effect, there's a 35% chance of seeing improvement this large or larger*
- ***Conclusion:*** *Weak evidence against the null hypothesis*
- ***Decision:*** *Fail to reject H_0 (method may not be effective)*

5.5 Determining the Direction: One-Tailed vs. Two-Tailed Tests

One of the most important decisions in hypothesis testing is determining the direction of your alternative hypothesis. This depends entirely on your research question.

Types of Hypothesis Tests

Two-Tailed	Tests if parameter differs from μ_0 in *either* direction
Left-Tailed	Tests if parameter is *less than* μ_0
Right-Tailed	Tests if parameter is *greater than* μ_0

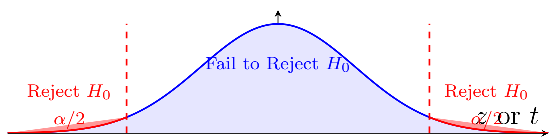
5.5.1 How to Choose the Direction

The key is to read your research question carefully and ask: "What am I trying to prove or demonstrate?"

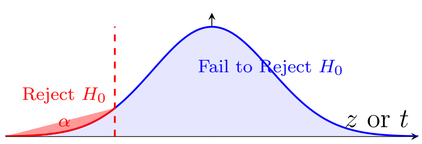
Research Question	**Null Hypothesis**	**Alternative**	**Test Type**
"Is the mean different from 20?"	$H_0 : \mu = 20$	$H_a : \mu \neq 20$	Two-tailed
"Is the machine under-filling?"	$H_0 : \mu = 5.0$	$H_a : \mu < 5.0$	Left-tailed
"Does the drug improve scores?"	$H_0 : \mu = 75$	$H_a : \mu > 75$	Right-tailed
"Is the new method better?"	$H_0 : \mu = 10$	$H_a : \mu > 10$	Right-tailed

5.5.2 Visual Guide to Critical Regions

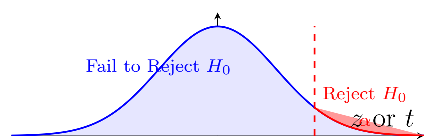
Two-Tailed Test: $H_a : \mu \neq \mu_0$



Left-Tailed: $H_a : \mu < \mu_0$



Right-Tailed: $H_a : \mu > \mu_0$



5.5.3 Common Keywords That Indicate Direction

Two-Tailed
$H_a : \mu \neq \mu_0$
different from
changed
not equal to
affects (either way)

Left-Tailed
$H_a : \mu < \mu_0$
less than
decreased
under-filling
worse than
below

Right-Tailed
$H_a : \mu > \mu_0$
greater than
increased
improved
better than
exceeds
above

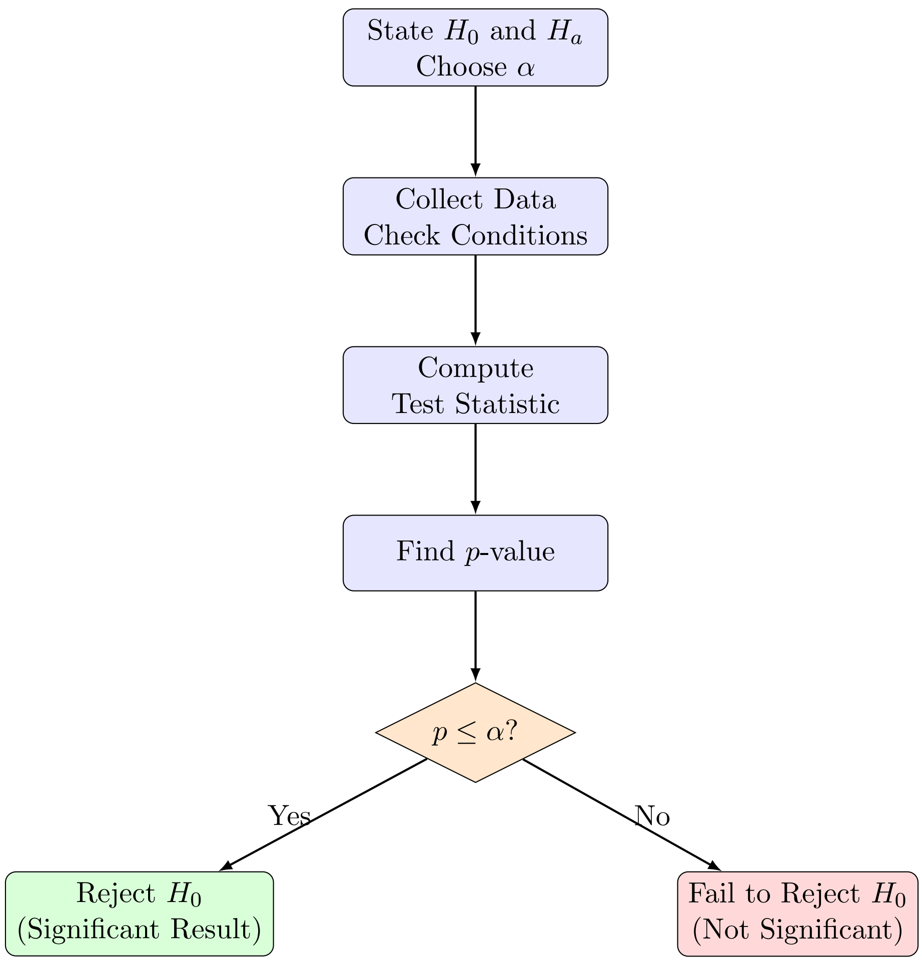
Decision Strategy

Ask yourself:

1. What am I trying to demonstrate or prove?
2. Am I looking for a change in a specific direction, or any change at all?
3. Does the context suggest I care about "worse" or "better" (one direction) or just "different" (either direction)?

Remember: The alternative hypothesis H_a represents what you're trying to provide evidence for!

5.6 Decision Flow Process



Interpretation: If the sample evidence would be rare under H_0 (small p-value), we deem the null implausible and *reject* it. Otherwise, we "fail to reject," acknowledging that the data are compatible with the status-quo claim.

6 Central Limit Theorem and Conditions

When we draw a random sample from a population, the *raw data* need not be normally distributed. The **Central Limit Theorem (CLT)** tells us that the *sampling distribution of the sample mean* $\bar{X}$ becomes approximately normal as sample size grows.

Central Limit Theorem (CLT)

When we collect a sufficiently large sample of n independent observations from a population with mean μ and standard deviation σ, the sampling distribution of $\bar{x}$ will be approximately normal with:

$$\text{Mean} = \mu \tag{1}$$

$$\text{Standard Error (SE)} = \frac{\sigma}{\sqrt{n}} \tag{2}$$

6.1 Required Conditions

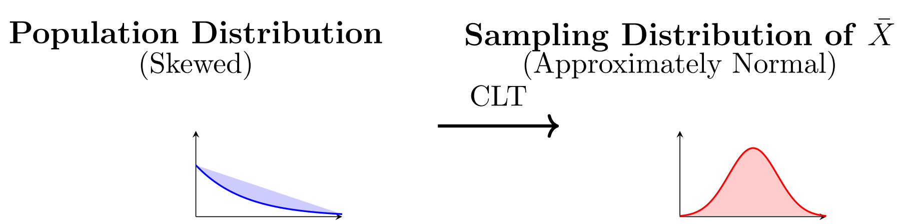
1. **Independence**: Sample observations must be independent. Most commonly satisfied when the sample is a simple random sample from the population.

2. **Normality**: When sample is small, we require that observations come from a normally distributed population. This condition can be relaxed for larger sample sizes.

Rules of Thumb: Normality Check

- $n < 30$: If sample size is less than 30 and there are no clear outliers, assume data come from a nearly normal distribution.

- $n \geq 30$: If sample size is at least 30 and there are no particularly extreme outliers, assume the sampling distribution of $\bar{x}$ is nearly normal, even if the underlying distribution is not.



6.2 The Problem: Unknown σ

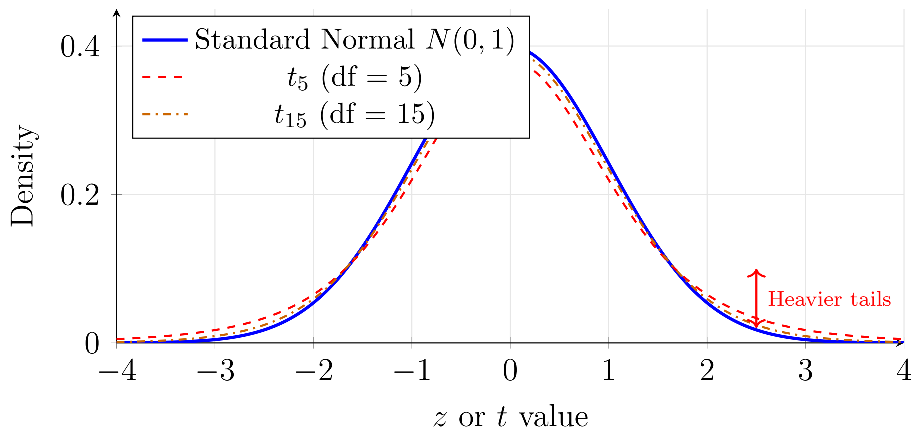
The standard error depends on the population standard deviation σ, which we rarely know. In practice, we substitute the sample standard deviation s:

$$\text{SE}(\bar{X}) \approx \frac{s}{\sqrt{n}}$$

This introduces additional uncertainty, especially when n is small. The remedy is to use the t-**distribution** instead of the normal distribution.

7 The t-Distribution

- A t-distribution is centered at 0 and controlled by degrees of freedom (df)
- For a sample mean based on n observations, we set $\text{df} = n - 1$
- As $\text{df} \to \infty$, the t-distribution approaches the standard normal
- For small df, it has visibly thicker tails



8 The One-Sample t-Test

8.1 When to Use

- Population standard deviation σ is *unknown* and sample sd s is used
- Sample is random; population is reasonably normal *or* $n \geq 30$ (CLT applies)

8.2 Test Statistic

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \quad \xrightarrow[H_0]{} \quad t_{n-1}$$

Example 8.1 (Protein Bar Analysis)**.** *A nutritionist claims protein bars contain 20g of protein on average. Ten bars are analyzed with results:*

$$19.1, 18.7, 21.0, 19.6, 20.3, 20.1, 18.9, 19.8, 20.4, 19.5$$

Test the claim at $\alpha = 0.05$.

Solution 8.1.1. ***Step 1: State Hypotheses***

$$H_0 : \mu = 20 \tag{3}$$
$$H_a : \mu \neq 20 \quad \text{(two-tailed test)} \tag{4}$$

Step 2: Check Conditions

- *Independence:* ✓ *(random sample)*
- *Normality:* ✓ *($n = 10 < 30$, no obvious outliers)*
- *Standard deviation:* ✓ *unknown → t-test*

Step 3: Calculate Sample Statistics

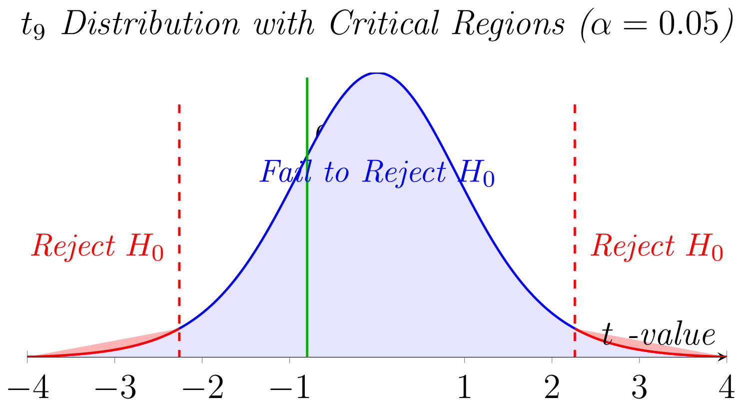
$$\bar{x} = \frac{19.1 + 18.7 + \cdots + 19.5}{10} = 19.74 \tag{5}$$
$$s = 0.63 \tag{6}$$
$$n = 10 \tag{7}$$

Step 4: Compute Test Statistic

$$t = \frac{19.74 - 20}{0.63/\sqrt{10}} = \frac{-0.26}{0.199} \approx -0.80$$

Step 5: Find Critical Values and p-value

- *Degrees of freedom: $df = n - 1 = 9$*
- *Critical values: $t_{0.025,9} = \pm 2.262$*
- *Critical region: $|t| > 2.262$*
- *p-value: $p = 2P(T_9 > |-0.80|) \approx 0.44$ (optional)*



Step 6: Make Decision *Since $|t| = 0.80 < 2.262$ and $p = 0.44 > 0.05$, we* ***fail to reject*** *H_0.*

Conclusion: *The data do not provide significant evidence that the true mean protein content differs from 20g at the $\alpha = 0.05$ level.*

9 The One-Sample z-Test

When the population standard deviation σ is known (or n is large enough that $s \approx \sigma$), we use the standard normal distribution.

9.1 Test Statistic

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \xrightarrow[H_0]{} \quad N(0,1)$$

Example 9.1 (Industrial Process Control)**.** *A filling machine is designed to fill bags with 5.0 kg of fertilizer. Historical data shows $\sigma = 0.20$ kg. A sample of $n = 50$ bags has mean fill 4.94 kg. At $\alpha = 0.05$, test if the process is under-filling.*

Solution 9.1.1. ***Step 1: State Hypotheses***

$$H_0 : \mu = 5.0 \quad \textit{(process is filling correctly)} \tag{8}$$

$$H_a : \mu < 5.0 \quad \textit{(process is under-filling)} \tag{9}$$

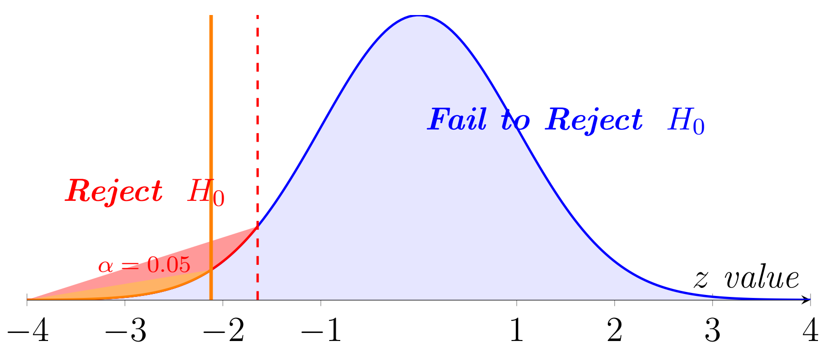
Step 2: Compute Test Statistic

$$z = \frac{4.94 - 5.0}{0.20/\sqrt{50}} = \frac{-0.06}{0.0283} \approx -2.12$$

Step 3: Find Critical Value and p-value

- *One-tailed test:* $z_{0.05} = -1.645$
- *p-value:* $P(Z < -2.12) \approx 0.017$ *(optional)*

Standard Normal Distribution - Left-Tailed Test ($\alpha = 0.05$)



Step 4: Decision *Since $z = -2.12 < -1.645$ and $p = 0.017 < 0.05$, we* ***reject*** *H_0.*
Conclusion: *There is significant evidence that the machine is under-filling bags.*

9.2 Using the Standard Normal (Z) Table

Understanding how to read the z-table is crucial for finding critical values and p-values in hypothesis testing.

How to Read the Z-Table

The standard normal table gives you the area to the **left** of a z-value under the standard normal curve.

- **Table Value** = $P(Z \leq z)$ = Area to the left of z
- **Rows**: First two digits of z-value (e.g., -2.1, 1.6)
- **Columns**: Third decimal place (e.g., 0.02, 0.05)

9.2.1 Sample Z-Table (Partial)

Standard Normal Table (Left-tail areas)

z	0.00	0.01	0.02	0.03	0.04	0.05
-2.2	0.0139	0.0136	**0.0132**	0.0129	0.0125	0.0122
-2.1	0.0179	0.0174	$\boxed{\textbf{0.0170}}$ $z = -2.12$ $p = 0.017$	0.0166	0.0162	0.0158
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202
-1.7	0.0446	0.0436	0.0427	0.0418	**0.0409**	$\boxed{\textbf{0.0401}}$ $z_{0.05} = -1.645$ Critical Value
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495

9.2.2 Step-by-Step Examples

Example 1: Finding p-value for z

Goal: Find $P(Z < -2.12)$
Steps:

1. Look up row for $z = -2.1$
2. Look up column for 0.02 (since $-2.12 = -2.1 + (-0.02)$)
3. Find intersection: $P(Z < -2.12) = 0.0170 \approx 0.017$

This matches our p-value from the fertilizer example!

Example 2: Finding critical value for α

Goal: Find z such that $P(Z < z) = 0.05$
Steps:

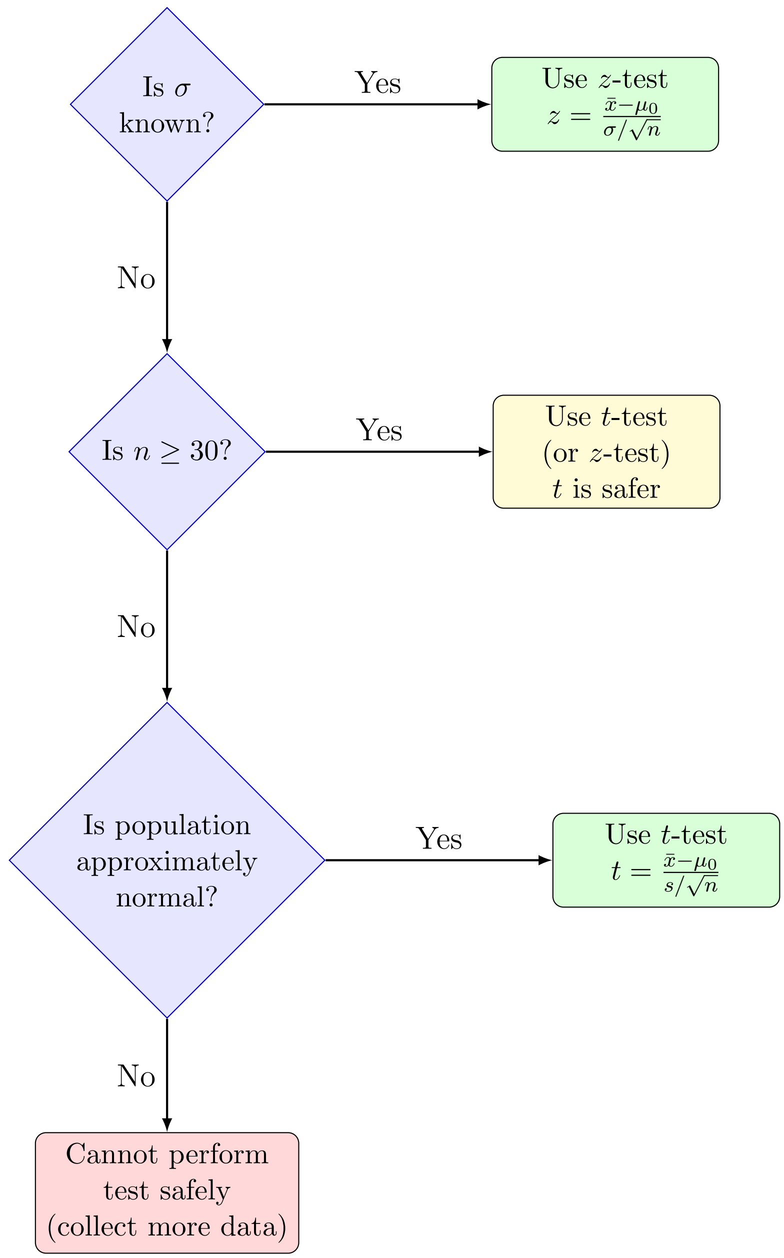
1. Look through the table body for value closest to 0.0500
2. Find 0.0495 at row $z = -1.6$, column 0.05
3. Critical value: $z_{0.05} = -1.645$

This is our critical value from the fertilizer example!

Key Points for Using Z-Tables

- **Always remember:** Table gives area to the *left*
- **For right-tail areas:** Use $P(Z > z) = 1 - P(Z < z)$
- **For two-tail tests:** Find area in one tail, then double it
- **Critical values:** Look up the α area in the table body, find corresponding z
- **P-values:** Look up your calculated z-statistic, read the probability

10 Decision Guide: z-test vs. t-test



11 Quick Reference Guide

Statistical Test Summary

Test Type	Test Statistic	Distribution	When to Use
One-sample z	$z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$N(0, 1)$	σ known
One-sample t	$t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$	t_{n-1}	σ unknown
One proportion	$z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$	$N(0, 1)$	$np \geq 10$, $n(1 - p) \geq 10$

11.1 Type I and Type II Errors

2***Decision**	**Reality**	
	H_0 True	H_0 False
Reject H_0	Type I Error (α)	Correct Decision (Power)
Fail to Reject H_0	Correct Decision	Type II Error (β)

- **Type I Error**: Rejecting a true null hypothesis (false positive)
- **Type II Error**: Failing to reject a false null hypothesis (false negative)
- **Power**: Probability of correctly rejecting a false null hypothesis ($1 - \beta$)

11.2 Key Reminders

Important Notes

1. Always check conditions before performing any test
2. A non-significant result does not prove H_0 is true
3. Statistical significance does not imply practical significance
4. The p-value is NOT the probability that H_0 is true
5. Always interpret results in the context of the problem

12 Z-test vs. Z-score: What's the Difference?

Common Student Question

Are a z-test and a z-score the same?

No, but they are closely related.

12.1 Z-score (Standard Score)

Z-score Definition

A **z-score** (also called a *standard score*) tells you how many standard deviations a data point is from the population mean. It standardizes individual data values.

$$z = \frac{x - \mu}{\sigma}$$

Components:

- x = individual data point
- μ = population mean
- σ = population standard deviation

12.2 Z-test (Hypothesis Test)

Z-test Definition

A **z-test** is a hypothesis test used to assess whether a sample mean differs significantly from a hypothesized population mean when the population standard deviation is known.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Components:

- $\bar{x}$ = sample mean
- μ_0 = hypothesized population mean (from H_0)
- σ = population standard deviation
- n = sample size

12.3 Visual Comparison

Z-Score

Purpose: Standardize one data point

Answers: "How unusual is this single value?"

Example: Student's test score compared to class average

Both use standardization →

Z-Test

Purpose: Test hypothesis about sample mean

Answers: "Is this sample mean significantly different?"

Example: Is machine filling bags correctly?

12.4 Summary Comparison

Side-by-Side Comparison

Aspect	Z-score	Z-test
Purpose	Standardize a single data point	Test sample mean vs. population mean
Input	One value x	Sample: $\bar{x}$, n; Population: μ_0, σ
Output	Standard score (number)	Test statistic $\to$ p-value $\to$ decision
Use Case	Compare individual to population	Hypothesis testing on sample means
Example	"John scored 85 on a test with $\mu = 75$, $\sigma = 10$. His z-score is 1.0"	"Sample of 50 bags has $\bar{x} = 4.94$ kg. Is $\mu < 5.0$ kg?"

Key Takeaway

- **Z-score**: Describes how far *one data point* is from the mean
- **Z-test**: Uses a z-score-like calculation to test whether a *sample mean* is significantly different from a hypothesized value
- Both involve standardization, but serve different statistical purposes!