

Statistics for Economists II

Lecture 10: Errors, Power, & Testing with Unknown Variance

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Type I and Type II Errors
- 2 Power of a Test
- 3 Hypothesis Testing with Unknown Variance: The t -Test
- 4 Testing for a Population Proportion
- 5 Statistical vs. Economic Significance
- 6 Practice Problems
- 7 Key Takeaways

Recap: The Two Kinds of Mistakes

Every hypothesis test can produce two types of errors:

	H_0 true	H_0 false
Reject H_0	Type I	Correct!
Fail to reject H_0	Correct!	Type II

- $\alpha = \mathbb{P}(\text{Type I}) = \mathbb{P}(\text{reject } H_0 \mid H_0 \text{ true})$
- $\beta = \mathbb{P}(\text{Type II}) = \mathbb{P}(\text{fail to reject } H_0 \mid H_0 \text{ false})$

- **Type I:** Convicting an innocent person (false alarm)
- **Type II:** Letting a guilty person go free (missed detection)

Both are bad—but in different ways!

The Only Free Lunch: Increase Sample Size

Key Insight

The **only way** to reduce both α and β simultaneously is to **increase the sample size n** .

Why does a larger n help?

- Larger $n \implies$ smaller standard error $\sigma/\sqrt{n}$
- Smaller SE $\implies$ sampling distributions of $\bar{X}$ under H_0 and H_1 **overlap less**
- Less overlap $\implies$ easier to tell the two hypotheses apart

Think of it as **improving the resolution of a microscope**. With more data, you can see finer differences between what the world looks like under H_0 versus H_1 . With a tiny sample, everything is blurry.

Calculating β for a Specific Alternative

Setup: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1 > \mu_0$, with known σ , significance level α .

Decision rule: Reject H_0 if $\bar{X} > c$, where $c = \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$.

Calculating β for a Specific Alternative

Setup: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1 > \mu_0$, with known σ , significance level α .

Decision rule: Reject H_0 if $\bar{X} > c$, where $c = \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$.

Computing β : If the *true* mean is μ_1 ,

$$\begin{aligned}\beta(\mu_1) &= \mathbb{P}(\text{fail to reject } H_0 \mid \mu = \mu_1) = \mathbb{P}(\bar{X} \leq c \mid \mu = \mu_1) \\ &= \mathbb{P}\left(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} \leq \frac{c - \mu_1}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha\right)\end{aligned}$$

β depends on **how far the true μ_1 is from μ_0** . The farther apart they are, the smaller β —it is easier to detect a big difference!

Example: Computing β

Problem: A pharmaceutical company claims a drug lowers cholesterol by $\mu_0 = 10$ mg/dL. We test $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$ at $\alpha = 0.05$, with $\sigma = 8$, $n = 64$.

What is β if the true effect is $\mu_1 = 12$ mg/dL?

Example: Computing β

Problem: A pharmaceutical company claims a drug lowers cholesterol by $\mu_0 = 10$ mg/dL. We test $H_0 : \mu = 10$ vs. $H_1 : \mu > 10$ at $\alpha = 0.05$, with $\sigma = 8$, $n = 64$.

What is β if the true effect is $\mu_1 = 12$ mg/dL?

Solution:

- Critical value: $c = 10 + 1.645 \cdot \frac{8}{\sqrt{64}} = 10 + 1.645 = 11.645$
- Under $\mu_1 = 12$: $Z = \frac{11.645 - 12}{8/\sqrt{64}} = \frac{-0.355}{1} = -0.355$
- $\beta = \Phi(-0.355) = 0.3613$

Interpretation

There is a **36.1% chance** we would fail to detect the real effect of 12 mg/dL. That is a high probability of missing a genuinely effective drug!

What Is the Power of a Test?

Definition

The **power** of a test against a specific alternative μ_1 is:

$$\text{Power}(\mu_1) = 1 - \beta(\mu_1) = \mathbb{P}(\text{reject } H_0 \mid \mu = \mu_1)$$

Power is the probability of **correctly detecting a real effect**. It measures the test's **ability to find what is actually there**.

Think of it as the **sensitivity** of your test. A test with low power is like a metal detector that beeps only half the time there is actual metal—unreliable!

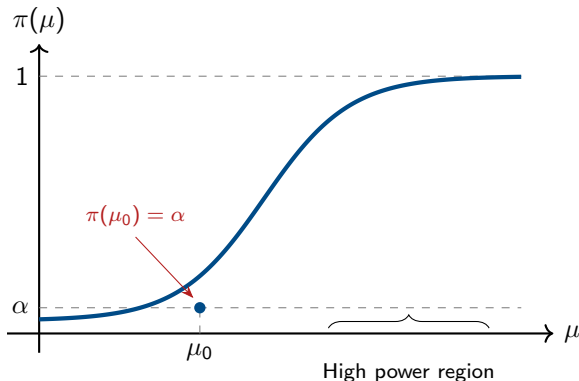
From our drug example: $\text{Power} = 1 - 0.3613 = 0.6387$.

We have only a 63.9% chance of detecting the real effect. Most researchers want power ≥ 0.80 .

The Power Function

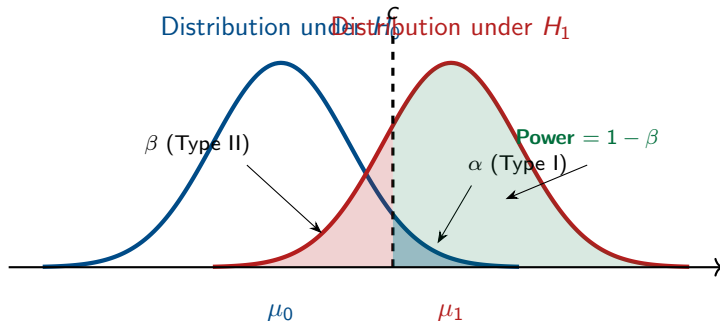
The **power function** $\pi(\mu)$ traces how power varies with the true parameter:

$$\pi(\mu) = \mathbb{P}(\text{reject } H_0 \mid \text{true mean is } \mu)$$



Properties: $\pi(\mu_0) = \alpha$ (the size of the test), and $\pi(\mu) \rightarrow 1$ as μ moves far from μ_0 .

Visualizing α , β , and Power



α = blue shaded area (false alarm rate). β = red shaded area (missed detection). **Power** = green shaded area (correct detection). Moving c left increases power but also increases α .

What Affects the Power of a Test?

Four factors determine power:

- ① **Sample size (n):** Larger $n \Rightarrow$ higher power
 - The distributions narrow and separate $\Rightarrow$ less overlap
- ② **Significance level (α):** Larger $\alpha \Rightarrow$ higher power
 - But at the cost of more Type I errors
- ③ **True parameter distance ($|\mu_1 - \mu_0|$):** Larger distance $\Rightarrow$ higher power
 - Big effects are easier to detect than small ones
- ④ **Population variance (σ^2):** Smaller variance $\Rightarrow$ higher power
 - Less noise in the data $\Rightarrow$ clearer signal

Summary Formula (one-sided, right-tail)

$$\text{Power} = 1 - \Phi\left(z_{\alpha} - \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}}\right)$$

Power and Sample Size: How Many Observations Do We Need?

We can **invert** the power formula to find the required n :

For a one-sided test at level α with desired power $1 - \beta$:

$$n = \left(\frac{(z_\alpha + z_\beta) \sigma}{\mu_1 - \mu_0} \right)^2$$

Power and Sample Size: How Many Observations Do We Need?

We can **invert** the power formula to find the required n :

For a one-sided test at level α with desired power $1 - \beta$:

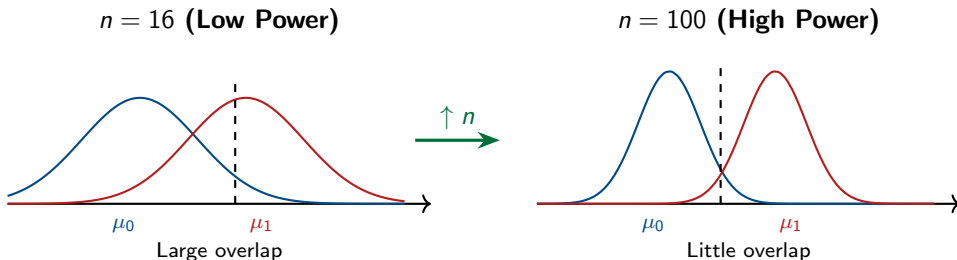
$$n = \left(\frac{(z_\alpha + z_\beta) \sigma}{\mu_1 - \mu_0} \right)^2$$

Example: Drug trial with $\sigma = 8$, $\mu_0 = 10$, $\mu_1 = 12$, $\alpha = 0.05$, desired power = 0.80.

$$\begin{aligned} n &= \left(\frac{(1.645 + 0.842) \times 8}{12 - 10} \right)^2 = \left(\frac{2.487 \times 8}{2} \right)^2 = \left(\frac{19.896}{2} \right)^2 \\ &= (9.948)^2 = 98.96 \approx \boxed{99} \end{aligned}$$

We need 99 patients (vs. our earlier 64) to have an 80% chance of detecting the 2 mg/dL effect. **Power analysis should be done before collecting data!**

How Power Changes with n : A Visual



As n increases, both distributions become **narrower** (SE shrinks), so they **separate**. This is why β drops and power rises—the two hypotheses become easier to tell apart.

The Problem: σ Is Usually Unknown

Reality Check

In virtually all real-world applications, we do **not** know the population standard deviation σ . We must estimate it from the sample using:

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

What changes?

- We replace σ with S in our test statistic
- But S is a **random variable**—it introduces additional uncertainty
- The resulting statistic no longer follows a normal distribution
- It follows a t -distribution with $n - 1$ degrees of freedom

Using S instead of σ is like using an estimated ruler whose markings are slightly uncertain.

The t -Test Statistic

From Z to T

Known σ : $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$

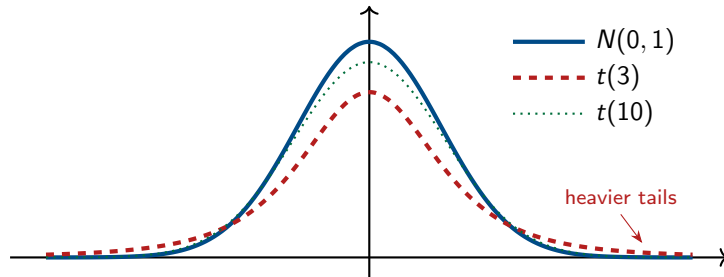
Unknown σ : $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n - 1)$

Key properties of the t -distribution:

- Symmetric and bell-shaped, like the normal
- **Heavier tails** than the normal (more probability in extremes)
- Indexed by **degrees of freedom** $\nu = n - 1$
- As $n \rightarrow \infty$, $t(n - 1) \rightarrow N(0, 1)$

The extra uncertainty from estimating σ makes extreme values **more likely**. The t -distribution accounts for this by having fatter tails. With large n , the estimation of σ becomes precise and $t \approx z$.

t -Distribution vs. Normal: Visual Comparison



Practical rule: For $n \geq 30$ (i.e., $\nu \geq 29$), $t(\nu)$ is very close to $N(0, 1)$. For small samples, the difference matters and we **must** use t -critical values.

Decision Rules for the t -Test

Test $H_0 : \mu = \mu_0$ using $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ with $\nu = n - 1$:

Alternative	Reject H_0 if	p -value
$H_1 : \mu > \mu_0$	$T > t_{\alpha, \nu}$	$\mathbb{P}(t_\nu > T_{\text{obs}})$
$H_1 : \mu < \mu_0$	$T < -t_{\alpha, \nu}$	$\mathbb{P}(t_\nu < T_{\text{obs}})$
$H_1 : \mu \neq \mu_0$	$ T > t_{\alpha/2, \nu}$	$2\mathbb{P}(t_\nu > T_{\text{obs}})$

Important Note

Critical values $t_{\alpha, \nu}$ come from the t -table (or software), **not** the standard normal table. For small n , these are noticeably larger than z_α .

Example: $t_{0.025, 5} = 2.571$ vs. $z_{0.025} = 1.960$. You need stronger evidence with small samples!

Example 5.3: Testing a Mean with Unknown σ

Problem: A government agency claims the average processing time for work permits is $\mu = 30$ days. A random sample of $n = 25$ applications yields $\bar{x} = 33.2$ days and $s = 6.5$ days. Test at $\alpha = 0.05$ whether the true mean exceeds 30 days.

Example 5.3: Testing a Mean with Unknown σ

Problem: A government agency claims the average processing time for work permits is $\mu = 30$ days. A random sample of $n = 25$ applications yields $\bar{x} = 33.2$ days and $s = 6.5$ days. Test at $\alpha = 0.05$ whether the true mean exceeds 30 days.

Step 1: Hypotheses. $H_0 : \mu = 30$ vs. $H_1 : \mu > 30$ (one-sided, right-tail)

Example 5.3: Testing a Mean with Unknown σ

Problem: A government agency claims the average processing time for work permits is $\mu = 30$ days. A random sample of $n = 25$ applications yields $\bar{x} = 33.2$ days and $s = 6.5$ days. Test at $\alpha = 0.05$ whether the true mean exceeds 30 days.

Step 1: Hypotheses. $H_0 : \mu = 30$ vs. $H_1 : \mu > 30$ (one-sided, right-tail)

Step 2: Test statistic.

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{33.2 - 30}{6.5/\sqrt{25}} = \frac{3.2}{1.3} = 2.462$$

Example 5.3: Testing a Mean with Unknown σ

Problem: A government agency claims the average processing time for work permits is $\mu = 30$ days. A random sample of $n = 25$ applications yields $\bar{x} = 33.2$ days and $s = 6.5$ days. Test at $\alpha = 0.05$ whether the true mean exceeds 30 days.

Step 1: Hypotheses. $H_0 : \mu = 30$ vs. $H_1 : \mu > 30$ (one-sided, right-tail)

Step 2: Test statistic.

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{33.2 - 30}{6.5/\sqrt{25}} = \frac{3.2}{1.3} = 2.462$$

Step 3: Critical value. $t_{0.05,24} = 1.711$ ($\nu = n - 1 = 24$)

Example 5.3: Testing a Mean with Unknown σ

Problem: A government agency claims the average processing time for work permits is $\mu = 30$ days. A random sample of $n = 25$ applications yields $\bar{x} = 33.2$ days and $s = 6.5$ days. Test at $\alpha = 0.05$ whether the true mean exceeds 30 days.

Step 1: Hypotheses. $H_0 : \mu = 30$ vs. $H_1 : \mu > 30$ (one-sided, right-tail)

Step 2: Test statistic.

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{33.2 - 30}{6.5/\sqrt{25}} = \frac{3.2}{1.3} = 2.462$$

Step 3: Critical value. $t_{0.05,24} = 1.711$ ($\nu = n - 1 = 24$)

Step 4: Decision. $T = 2.462 > 1.711 = t_{0.05,24} \implies$ **Reject H_0 .**

At the 5% significance level, there is sufficient evidence that the average processing time exceeds 30 days.

Example 5.3: Discussion

What if we had used a z -test instead?

- $z_{0.05} = 1.645$
- We would *still* reject, but the t -test is the **correct** procedure here
- With $n = 25$ and unknown σ , the t -distribution accounts for the extra uncertainty

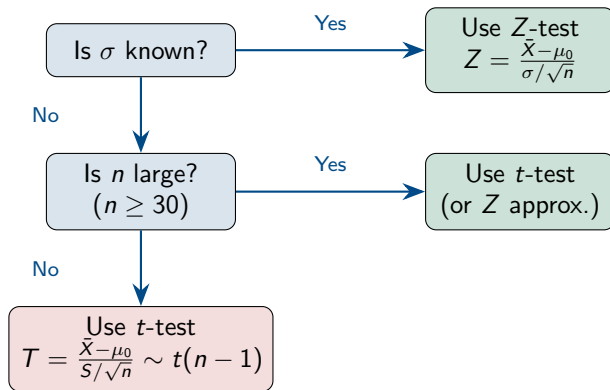
p -value: $\mathbb{P}(t_{24} > 2.462) \approx 0.011$

The p -value is small—strong evidence against H_0 .

In policy evaluation, this test could inform decisions about administrative reform. If processing times are significantly above target, resources may need reallocation.

But always ask: is the 3.2-day difference **practically significant**?
More on this later!

When Can We Use z Instead of t ?



Practical Advice

In modern statistics, **always use the t -test** when σ is unknown. It is exact for normal populations and approximately correct otherwise. The z -approximation for large n is a convenient shortcut, but the t -test never hurts.

Testing a Proportion: Setup

When the parameter of interest is a population proportion p :

Test Statistic for Proportions

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \underset{\sim}{\text{approx}} N(0, 1)$$

where $\hat{p} = X/n$ is the sample proportion and p_0 is the hypothesized value.

Why p_0 in the denominator (not $\hat{p}$)?

- Under H_0 , the true proportion *is* p_0
- So we use p_0 to compute the standard error under the null hypothesis
- Valid when $np_0 \geq 5$ and $n(1 - p_0) \geq 5$ (normal approximation condition)

Example 5.4: Minimum Wage and Employment

Problem: A policymaker claims that less than 20% of small businesses reduced employment after a minimum wage increase. In a survey of $n = 200$ firms, $\hat{p} = 34/200 = 0.17$ reported reductions. Test at $\alpha = 0.05$.

Example 5.4: Minimum Wage and Employment

Problem: A policymaker claims that less than 20% of small businesses reduced employment after a minimum wage increase. In a survey of $n = 200$ firms, $\hat{p} = 34/200 = 0.17$ reported reductions. Test at $\alpha = 0.05$.

Step 1: $H_0 : p = 0.20$ vs. $H_1 : p < 0.20$ (one-sided, left-tail)

Example 5.4: Minimum Wage and Employment

Problem: A policymaker claims that less than 20% of small businesses reduced employment after a minimum wage increase. In a survey of $n = 200$ firms, $\hat{p} = 34/200 = 0.17$ reported reductions. Test at $\alpha = 0.05$.

Step 1: $H_0 : p = 0.20$ vs. $H_1 : p < 0.20$ (one-sided, left-tail)

Step 2: Check conditions: $np_0 = 200 \times 0.20 = 40 \geq 5 \checkmark$ $n(1 - p_0) = 160 \geq 5 \checkmark$

Example 5.4: Minimum Wage and Employment

Problem: A policymaker claims that less than 20% of small businesses reduced employment after a minimum wage increase. In a survey of $n = 200$ firms, $\hat{p} = 34/200 = 0.17$ reported reductions. Test at $\alpha = 0.05$.

Step 1: $H_0 : p = 0.20$ vs. $H_1 : p < 0.20$ (one-sided, left-tail)

Step 2: Check conditions: $np_0 = 200 \times 0.20 = 40 \geq 5 \checkmark$ $n(1 - p_0) = 160 \geq 5 \checkmark$

Step 3: Test statistic:

$$Z = \frac{0.17 - 0.20}{\sqrt{\frac{0.20 \times 0.80}{200}}} = \frac{-0.03}{\sqrt{0.0008}} = \frac{-0.03}{0.02828} = -1.061$$

Example 5.4: Minimum Wage and Employment

Problem: A policymaker claims that less than 20% of small businesses reduced employment after a minimum wage increase. In a survey of $n = 200$ firms, $\hat{p} = 34/200 = 0.17$ reported reductions. Test at $\alpha = 0.05$.

Step 1: $H_0 : p = 0.20$ vs. $H_1 : p < 0.20$ (one-sided, left-tail)

Step 2: Check conditions: $np_0 = 200 \times 0.20 = 40 \geq 5 \checkmark$ $n(1 - p_0) = 160 \geq 5 \checkmark$

Step 3: Test statistic:

$$Z = \frac{0.17 - 0.20}{\sqrt{\frac{0.20 \times 0.80}{200}}} = \frac{-0.03}{\sqrt{0.0008}} = \frac{-0.03}{0.02828} = -1.061$$

Step 4: Critical value: $-z_{0.05} = -1.645$.

Since $-1.061 > -1.645$, we **fail to reject** H_0 .

At the 5% level, there is insufficient evidence that fewer than 20% of firms reduced employment.

Example 5.4: Interpretation

p -value: $\mathbb{P}(Z < -1.061) = 0.1443$

This is far above $\alpha = 0.05$ —weak evidence against H_0 .

What does “fail to reject” mean here?

- We **cannot conclude** that fewer than 20% of firms cut jobs
- This does *not* prove that $p = 0.20$
- The sample of 200 may lack the **power** to detect a difference of 3 percentage points

This is a classic policy-relevant test. The minimum wage debate hinges on whether employment effects are “significant.”

A power analysis might reveal that we need $n > 1,000$ firms to reliably detect a 3-percentage-point difference. The failure to reject may reflect **low power**, not absence of an effect.

A Large Sample Will Reject *Any* Null Hypothesis

For any fixed $\mu_1 \neq \mu_0$, no matter how small the difference:

$$\text{Power} = 1 - \Phi\left(z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma/\sqrt{n}}\right) \xrightarrow{n \rightarrow \infty} 1$$

With enough data, you will **always** find a “statistically significant” difference!

A Large Sample Will Reject *Any* Null Hypothesis

For any fixed $\mu_1 \neq \mu_0$, no matter how small the difference:

$$\text{Power} = 1 - \Phi\left(z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma/\sqrt{n}}\right) \xrightarrow{n \rightarrow \infty} 1$$

With enough data, you will **always** find a “statistically significant” difference!

A study with $n = 100,000$ might find that a training program increases wages by \$0.50/week with $p < 0.001$. Highly significant! But is it **economically meaningful**?

Statistical significance tells you the effect is **not zero**. It does not tell you the effect is **important**.

Statistical Significance $\neq$ Economic Significance

Statistical Significance

- Can we rule out that the effect is zero?
- Depends on n , α , and noise
- A tiny effect with huge n will be “significant”

Economic (Practical) Significance

- Is the effect **large enough to matter**?
- Depends on context, policy costs, and stakes
- A large effect with small n may be “insignificant”

Best practice: Always report:

- 1 The **point estimate** (how big is the effect?)
- 2 The **confidence interval** (how precise is our estimate?)
- 3 The **p -value** (how strong is the evidence against H_0 ?)

Effect size measures how large an effect is relative to the variability. Common measures: Cohen's $d = \frac{|\bar{x} - \mu_0|}{s}$. This is independent of sample size and captures practical importance.

Practice Problem 1: Power Calculation

Problem: An economist tests whether a tax reform increased average household savings beyond \$500/month. She tests $H_0 : \mu = 500$ vs. $H_1 : \mu > 500$ at $\alpha = 0.05$ with $\sigma = 120$ and $n = 36$.

- (a) Find β and power if the true mean is $\mu_1 = 540$.
- (b) How large must n be for 90% power against $\mu_1 = 540$?

Practice Problem 1: Power Calculation

Problem: An economist tests whether a tax reform increased average household savings beyond \$500/month. She tests $H_0 : \mu = 500$ vs. $H_1 : \mu > 500$ at $\alpha = 0.05$ with $\sigma = 120$ and $n = 36$.

- (a) Find β and power if the true mean is $\mu_1 = 540$.
- (b) How large must n be for 90% power against $\mu_1 = 540$?

Solution (a):

- Critical value: $c = 500 + 1.645 \cdot \frac{120}{\sqrt{36}} = 500 + 1.645 \times 20 = 532.9$
- Under $\mu_1 = 540$: $Z = \frac{532.9 - 540}{120/\sqrt{36}} = \frac{-7.1}{20} = -0.355$
- $\beta = \Phi(-0.355) = 0.3613$
- Power = $1 - 0.3613 = 0.6387$ (only 63.9%)

Practice Problem 1 (continued)

Solution (b): For power = 0.90 ($\beta = 0.10$), $z_\beta = 1.282$.

$$\begin{aligned} n &= \left(\frac{(z_\alpha + z_\beta) \sigma}{\mu_1 - \mu_0} \right)^2 = \left(\frac{(1.645 + 1.282) \times 120}{540 - 500} \right)^2 \\ &= \left(\frac{2.927 \times 120}{40} \right)^2 = \left(\frac{351.24}{40} \right)^2 = (8.781)^2 = 77.1 \end{aligned}$$

$$n \geq 78$$

Going from 64% power to 90% power requires more than doubling our sample from 36 to 78. Power does not scale linearly with n —it has diminishing returns. But the investment is worthwhile: with 78 observations, we have a 90% chance of detecting the \$40 increase.

Practice Problem 2: t -Test

Problem: A study examines whether a microfinance program increases monthly income beyond \$350 for participants. A sample of $n = 20$ borrowers yields $\bar{x} = 372$ and $s = 45$. Test at $\alpha = 0.01$.

Practice Problem 2: t -Test

Problem: A study examines whether a microfinance program increases monthly income beyond \$350 for participants. A sample of $n = 20$ borrowers yields $\bar{x} = 372$ and $s = 45$. Test at $\alpha = 0.01$.

Solution:

$$H_0 : \mu = 350 \quad \text{vs.} \quad H_1 : \mu > 350$$

$$T = \frac{372 - 350}{45/\sqrt{20}} = \frac{22}{10.062} = 2.186$$

Critical value: $t_{0.01,19} = 2.539$

Since $T = 2.186 < 2.539$, we **fail to reject** H_0 at $\alpha = 0.01$.

Note

At $\alpha = 0.05$, $t_{0.05,19} = 1.729$, and we *would* reject. The conclusion depends on the chosen significance level! The p -value ≈ 0.021 gives the full picture.

Practice Problem 3: Proportion Test with Power

Problem: Before an election, a pollster tests whether support for a candidate exceeds 50%. A sample of $n = 400$ voters shows $\hat{p} = 0.54$.

- (a) Test $H_0 : p = 0.50$ vs. $H_1 : p > 0.50$ at $\alpha = 0.05$.
- (b) What is the power of this test if the true proportion is $p_1 = 0.55$?

Practice Problem 3: Proportion Test with Power

Problem: Before an election, a pollster tests whether support for a candidate exceeds 50%. A sample of $n = 400$ voters shows $\hat{p} = 0.54$.

- (a) Test $H_0 : p = 0.50$ vs. $H_1 : p > 0.50$ at $\alpha = 0.05$.
- (b) What is the power of this test if the true proportion is $p_1 = 0.55$?

Solution (a):

$$Z = \frac{0.54 - 0.50}{\sqrt{\frac{0.50 \times 0.50}{400}}} = \frac{0.04}{0.025} = 1.60$$

$z_{0.05} = 1.645$. Since $1.60 < 1.645$, **fail to reject H_0** . (p -value = 0.0548)

Tantalizingly close! The evidence is suggestive but not quite significant at 5%.

Practice Problem 3 (continued)

Solution (b): Power against $p_1 = 0.55$.

Critical value for $\hat{p}$: $c = 0.50 + 1.645 \times 0.025 = 0.5411$

Under $p_1 = 0.55$, $SE_1 = \sqrt{\frac{0.55 \times 0.45}{400}} = 0.02487$

$$\begin{aligned}\text{Power} &= \mathbb{P}(\hat{p} > 0.5411 \mid p = 0.55) = \mathbb{P}\left(Z > \frac{0.5411 - 0.55}{0.02487}\right) \\ &= \mathbb{P}(Z > -0.358) = \Phi(0.358) = 0.6398\end{aligned}$$

With 400 voters, we have only a 64% chance of detecting a true support level of 55%. This explains why close elections are so hard to call from polls. To achieve 80% power, we would need:

$$n = \left(\frac{(1.645 + 0.842) \times \sqrt{0.5 \times 0.5}}{0.55 - 0.50} \right)^2 \approx 620 \text{ voters}$$

Practice Problem 4: Choosing the Right Test

Problem: An environmental economist measures pollution levels at 12 industrial sites. The regulatory limit is 50 ppm. The sample gives $\bar{x} = 54.3$ ppm, $s = 8.1$ ppm. Is there evidence that the average exceeds the limit? ($\alpha = 0.05$)

Practice Problem 4: Choosing the Right Test

Problem: An environmental economist measures pollution levels at 12 industrial sites. The regulatory limit is 50 ppm. The sample gives $\bar{x} = 54.3$ ppm, $s = 8.1$ ppm. Is there evidence that the average exceeds the limit? ($\alpha = 0.05$)

Solution: Since σ is unknown and $n = 12$ is small, use the t -test.

$$T = \frac{54.3 - 50}{8.1/\sqrt{12}} = \frac{4.3}{2.338} = 1.839$$

$$t_{0.05,11} = 1.796$$

Since $T = 1.839 > 1.796$, **reject** H_0 . There is evidence at the 5% level that average pollution exceeds the limit.

Policy Implication

This result justifies regulatory action. Note: if we had mistakenly used $z_{0.05} = 1.645$, we would also reject—but the t -test is the correct procedure for $n = 12$ with unknown σ . The margin is slim ($p \approx 0.047$).

Summary

- ① **α - β tradeoff:** For fixed n , reducing Type I error increases Type II error. The only way to reduce both is to **increase sample size**.
- ② **Power** = $1 - \beta$ measures the test's ability to detect a real effect. Depends on n , α , $|\mu_1 - \mu_0|$, and σ^2 .
- ③ **Power analysis** should be conducted *before* data collection to determine the required sample size.
- ④ **Unknown σ :** Use the t -test with $\nu = n - 1$ degrees of freedom. Critical values are larger than z , especially for small n .
- ⑤ **Proportion tests** use the z -statistic with p_0 in the standard error.
- ⑥ **Statistical $\neq$ economic significance.** Large samples can make trivial effects significant. Always report effect sizes and confidence intervals alongside p -values.

Common Pitfalls to Avoid

Mistakes Students Make

- 1 Using z-critical values when σ is unknown and n is small
- 2 Confusing “fail to reject H_0 ” with “ H_0 is true”
- 3 Ignoring power—a non-significant result may reflect low power, not absence of effect
- 4 Treating $p < 0.05$ as a magic threshold— $p = 0.049$ and $p = 0.051$ are practically identical
- 5 Forgetting that the standard error in a proportion test uses p_0 , not $\hat{p}$

A wise statistician once said: “Absence of evidence is not evidence of absence.” When you fail to reject H_0 , check the power. If power is low, the test simply could not have detected the effect even if it existed.

Further Reading

- Wackerly, Mendenhall & Scheaffer, Ch. 10: Hypothesis Testing
- Hogg, Tanis & Zimmerman, Ch. 6: Point Estimation and Tests of Hypotheses
- Devore & Berk, Ch. 8: Tests of Hypotheses Based on a Single Sample
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*
- Wasserstein & Lazar (2016). “The ASA Statement on p -Values”—*The American Statistician*

Coming Up Next Week

Two-sample tests: comparing means and proportions between two populations, paired vs. independent samples, and the pooled t -test.