Economics 103 – Statistics for Economists

Minsu Chang

University of Pennsylvania

Lecture 22

Last Time

Walked through steps of hypothesis testing in a simple example.

Today

- Relationship between hypothesis testing and CIs
- More examples of hypothesis tests

Relationship between CI and Two-Sided Test

- There is a *very close* relationship between CIs and hypothesis tests against a two-sided alternative.
- I'll illustrate this using a generic version of the example from last class but the relationship holds *in general*.

Relationship between CI and Two-sided Test

Suppose $X_1, \ldots, X_n \sim \text{iid } N(\mu, \sigma^2)$

Test $H_0\colon \mu = \mu_0$ vs. $H_1\colon \mu \neq \mu_0$ at significance level α

- Test Statistic: $T_n = \sqrt{n}(\bar{X}_n - \mu_0)/S \sim t(n-1)$ under H_0
- Decision Rule: Reject H_0 if $|T_n| > \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)$

Relationship between CI and Two-sided Test

Suppose $X_1, \ldots, X_n \sim$ iid $N(\mu, \sigma^2)$

Test $H_0\colon \mu = \mu_0$ vs. $H_1\colon \mu \neq \mu_0$ at significance level α

- Test Statistic: $T_n = \sqrt{n}(\bar{X}_n - \mu_0)/S \sim t(n-1)$ under H_0
- Decision Rule: Reject H_0 if $|T_n| > \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)$

$100 \times (1-\alpha)\%$ CI for μ

$$\bar{X}_n \pm \texttt{qt}(1-\alpha/2, \texttt{df} = n-1)\frac{S}{\sqrt{n}}$$

Relationship between CI and Two-sided Test

$c = \mathtt{qt}(1 - \alpha/2, \mathtt{df} = n - 1)$

Decision Rule: Reject H_0 if

$$\left| \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \right| > c \quad \Longleftrightarrow$$

Relationship between CI and Two-sided Test

$c = \texttt{qt}(1 - \alpha/2, \texttt{df} = n - 1)$

Decision Rule: Reject H_0 if

$$\left|\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}\right| > c \quad \iff \quad \left(\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} > c \quad \text{OR} \quad \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} < -c\right)$$

Relationship between CI and Two-sided Test

$c = \texttt{qt}(1 - \alpha/2, \texttt{df} = n - 1)$

Decision Rule: Reject H_0 if

$$\left|\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}}\right| > c \quad \iff \quad \left(\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} > c \quad \text{OR} \quad \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} < -c\right)$$

Equivalent to: *Don't Reject* H_0 provided

$$-c \leq \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \leq c$$

Relationship between CI and Two-sided Test

$c = \mathtt{qt}(1 - \alpha/2, \mathtt{df} = n - 1)$

Decision Rule: Reject H_0 if

$$\left| \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \right| > c \quad \iff \quad \left(\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} > c \quad \text{OR} \quad \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} < -c \right)$$

Equivalent to: *Don't Reject* H_0 provided

$$-c \leq \frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \leq c$$

$$\bar{X}_n - c \times \frac{S}{\sqrt{n}} \leq \mu_0 \leq \bar{X}_n + c \times \frac{S}{\sqrt{n}}$$

What does this mean?

Two-sided Test $\iff$ Checking if $\mu_0 \in$ CI

A two-sided test of $H_0\colon \mu = \mu_0$ against $H_1\colon \mu \neq \mu_0$ at significance level α is equivalent to checking whether μ_0 lies inside the corresponding $100 \times (1 - \alpha)\%$ confidence interval for μ.

What does this mean?

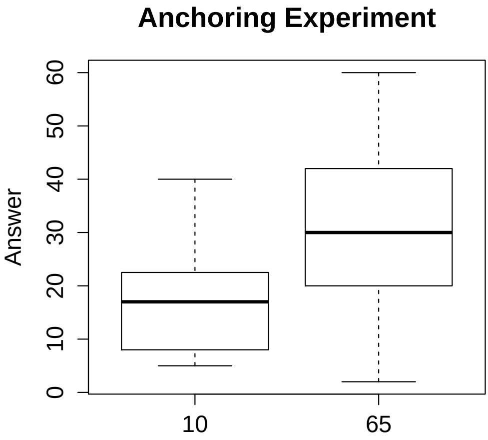
Two-sided Test $\iff$ Checking if $\mu_0 \in$ CI

A two-sided test of $H_0\colon \mu = \mu_0$ against $H_1\colon \mu \neq \mu_0$ at significance level α is equivalent to checking whether μ_0 lies inside the corresponding $100 \times (1 - \alpha)\%$ confidence interval for μ.

"Inverting" Two-sided Test to get a CI

Collect all the values μ_0 such that we cannot reject $H_0\colon \mu = \mu_0$ against the two-sided alternative. The result is *precisely* a $100 \times (1 - \alpha)\%$ CI for μ.

The Anchoring Experiment



The Anchoring Experiment

Shown a “random” number and then asked what proportion of UN member states are located in Africa.

“Hi” Group – Shown 65 ($n_{Hi} = 46$)

Sample Mean: 30.7, Sample Variance: 253

“Lo” Group – Shown 10 ($n_{Lo} = 43$)

Sample Mean: 17.1, Sample Variance: 86

Fairly large samples here, so we’ll proceed via the CLT…

In words, what is our null hypothesis?

(a) There is a *positive* anchoring effect: seeing a higher random number makes people report a higher answer.

(b) There is a *negative* anchoring effect: seeing a lower random number makes people report a lower answer.

(c) There *is* an anchoring effect: it could be positive or negative.

(d) There is *no* anchoring effect: people aren't influenced by seeing a random number before answering.

In symbols, what is our null hypothesis?

(a) $\mu_{Lo} < \mu_{Hi}$

(b) $\mu_{Lo} = \mu_{Hi}$

(c) $\mu_{Lo} > \mu_{Hi}$

(d) $\mu_{Lo} \neq \mu_{Hi}$

In symbols, what is our null hypothesis?

(a) $\mu_{Lo} < \mu_{Hi}$

(b) $\mu_{Lo} = \mu_{Hi}$

(c) $\mu_{Lo} > \mu_{Hi}$

(d) $\mu_{Lo} \neq \mu_{Hi}$

$\mu_{Lo} = \mu_{Hi}$ is *equivalent to* $\mu_{Hi} - \mu_{Lo} = 0$!

Anchoring Experiment

Under the null, what should we expect to be true about the values taken on by $\bar{X}_{Lo}$ and $\bar{X}_{Hi}$?

(a) They should be similar in value.

(b) $\bar{X}_{Lo}$ should be the smaller of the two.

(c) $\bar{X}_{Hi}$ should be the smaller of the two.

(d) They should be different. We don't know which will be larger.

What is our Test Statistic?

Sampling Distribution

$$\frac{(\bar{X}_{Hi} - \bar{X}_{Lo}) - (\mu_{Hi} - \mu_{Lo})}{\sqrt{\frac{S^2_{Hi}}{n_{Hi}} + \frac{S^2_{Lo}}{n_{Lo}}}} \approx N(0,1)$$

Test Statistic: Impose the Null

Under $H_0\colon \mu_{Lo} = \mu_{Hi}$

$$T_n = \frac{\bar{X}_{Hi} - \bar{X}_{Lo}}{\sqrt{\frac{S^2_{Hi}}{n_{Hi}} + \frac{S^2_{Lo}}{n_{Lo}}}} \approx N(0,1)$$

What is our Test Statistic?

$\bar{X}_{Hi} = 30.7,\ s^2_{Hi} = 253,\ n_{Hi} = 46$

$\bar{X}_{Lo} = 17.1,\ s^2_{Lo} = 86,\ n_{Lo} = 43$

Under $H_0: \mu_{Lo} = \mu_{Hi}$

$$T_n = \frac{\bar{X}_{Hi} - \bar{X}_{Lo}}{\sqrt{\frac{S^2_{Hi}}{n_{Hi}} + \frac{S^2_{Lo}}{n_{Lo}}}} \approx N(0,1)$$

Plugging in Our Data

$$T_n = \frac{\bar{X}_{Hi} - \bar{X}_{Lo}}{\sqrt{\frac{S^2_{Hi}}{n_{Hi}} + \frac{S^2_{Lo}}{n_{Lo}}}} \approx 5$$

Anchoring Experiment Example

Approximately what critical value should we use to test $H_0\colon \mu_{Lo} = \mu_{Hi}$ against the two-sided alternative at the 5% significance level?

Anchoring Experiment Example

Approximately what critical value should we use to test $H_0\colon \mu_{Lo} = \mu_{Hi}$ against the two-sided alternative at the 5% significance level?

α	0.10	0.05	0.01
`qnorm(1 −` α`)`	1.28	1.64	2.33
`qnorm(1 −` $\alpha/2$`)`	1.64	1.96	2.58

... Approximately 2

Anchoring Experiment Example

Which of these commands would give us the p-value of our test of $H_0: \mu_{Lo} = \mu_{Hi}$ against $H_1: \mu_{Lo} < \mu_{Hi}$ at significance level α?

(a) `qnorm(`$1 - \alpha$`)`

(b) `qnorm(`$1 - \alpha/2$`)`

(c) `1 - pnorm(5)`

(d) `2 * (1 - pnorm(5))`

P-values for $H_0: \mu_{Lo} = \mu_{Hi}$

We plug in the value of the test statistic that we observed: 5

Against $H_1: \mu_{Lo} < \mu_{Hi}$

`1 - pnorm(5)` < 0.0000

Against $H_1: \mu_{Lo} \neq \mu_{Hi}$

`2 * (1 - pnorm(5))` < 0.0000

If the null is true (the two population means are equal) it would be extremely unlikely to observe a test statistic as large as this!

What should we conclude?

Which Exam is Harder?

Student	Exam 1	Exam 2	Difference
1	57.1	60.7	3.6
⋮	⋮	⋮	⋮
71	78.6	82.9	4.3
Sample Mean:	79.6	81.4	1.8
Sample Var.	117	151	124
Sample Corr.	0.54		

Again, large sample size here so we'll use CLT.

One-Sample Hypothesis Test Using Differences

$\boxed{\text{Let } D_i = X_i - Y_i \text{ be (Midterm 2 Score - Midterm 1 Score) for student } i}$

Null Hypothesis

$H_0: \mu_1 = \mu_2$, i.e. both exams were of the same difficulty

Two-Sided Alternative

$H_1: \mu_1 \neq \mu_2$, i.e. one exam was harder than the other

One-Sided Alternative

$H_1: \mu_2 > \mu_1$, i.e. the second exam was easier

Decision Rules

$\boxed{\text{Let } D_i = X_i - Y_i \text{ be (Midterm 2 Score - Midterm 1 Score) for student } i}$

Test Statistic

$$\frac{\bar{D}_n}{\widehat{SE}(\bar{D}_n)} = \frac{1.8}{\sqrt{124/71}} \approx 1.36$$

Two-Sided Alternative

Reject $H_0\colon \mu_1 = \mu_2$ in favor of $H_1\colon \mu_1 \neq \mu_2$ if $|\bar{D}_n|$ is sufficiently large.

One-Sided Alternative

Reject $H_0\colon \mu_1 = \mu_2$ in favor of $H_1\colon \mu_2 > \mu_1$ if $\bar{D}_n$ is sufficiently large.

Reject against *Two-sided* Alternative with $\alpha = 0.1$?

$$\boxed{\frac{\bar{D}_n}{\widehat{SE}(\bar{D}_n)} = \frac{1.8}{\sqrt{124/71}} \approx 1.36}$$

α	0.10	0.05	0.01
$\texttt{qnorm}(1-\alpha)$	1.28	1.64	2.33
$\texttt{qnorm}(1-\alpha/2)$	1.64	1.96	2.58

(a) Reject

(b) Fail to Reject

(c) Not Sure

Reject against *One-sided* Alternative with $\alpha = 0.1$?

$$\boxed{\frac{\bar{D}_n}{\widehat{SE}(\bar{D}_n)} = \frac{1.8}{\sqrt{124/71}} \approx 1.36}$$

α	0.10	0.05	0.01
`qnorm(1` $-\ \alpha$`)`	1.28	1.64	2.33
`qnorm(1` $-\ \alpha/2$`)`	1.64	1.96	2.58

(a) Reject

(b) Fail to Reject

(c) Not Sure

P-Values for the Test of $H_0: \mu_1 = \mu_2$

$$\boxed{\frac{\bar{D}_n}{\widehat{SE}(\bar{D}_n)} = \frac{1.8}{\sqrt{124/71}} \approx 1.36}$$

One-Sided $H_1: \mu_2 > \mu_1$

`1 - pnorm(1.36)` $=$ `pnorm(-1.36)` ≈ 0.09

Two-Sided $H_1: \mu_1 \neq \mu_2$

`2 * (1 - pnorm(1.36))` $=$ `2 * pnorm(-1.36)` ≈ 0.18

Tests for Proportions

Basic Idea

The population *can't be* normal (it's Bernoulli) so we use the CLT to get approximate sampling distributions (c.f. Lecture 18).

But there's a small twist!

Bernoulli RV only has a *single* unknown parameter $\implies$ we know *more* about the population under H_0 in a proportions problem than in the other testing examples we've examined...

For best results, always *fully* impose the null.

Tests for Proportions: One-Sample Example

From Pew Polling Data

54% of a random sample of 771 registered voters correctly identified 2012 presidential candidate Mitt Romney as Pro-Life.

Sampling Model

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

Sample Statistic

Sample Proportion: $\widehat{p} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

$\boxed{\text{Suppose I wanted to test } H_0: p = 0.5}$

Tests for Proportions: One Sample Example

Under $H_0: p = 0.5$ what is the standard error of $\widehat{p}$?

(a) 1

(b) $\sqrt{\widehat{p}(1-\widehat{p})/n}$

(c) $\sigma/\sqrt{n}$

(d) $1/(2\sqrt{n})$

(e) $p(1-p)$

Tests for Proportions: One Sample Example

Under $H_0: p = 0.5$ what is the standard error of $\widehat{p}$?

(a) 1

(b) $\sqrt{\widehat{p}(1-\widehat{p})/n}$

(c) $\sigma/\sqrt{n}$

(d) $1/(2\sqrt{n})$

(e) $p(1-p)$

$p = 0.5 \implies \sqrt{0.5(1-0.5)/n} = 1/(2\sqrt{n})$

Under the null we know the SE! Don't have to estimate it!

One-Sample Test for a Population Proportion

Sampling Model

$X_1, \ldots, X_n \sim$ iid Bernoulli(p)

Null Hypothesis

$H_0\colon p =$ Known Constant p_0

Test Statistic

$$T_n = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \approx N(0,1) \text{ under } H_0 \text{ provided } n \text{ is large}$$

One-Sample Example H_0: $p = 0.5$

54% of a random sample of 771 registered voters knew Mitt Romney is Pro-Life.

$$T_n = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = 2\sqrt{771}(0.54 - 0.5)$$

$$= 0.08 \times \sqrt{771} \approx 2.2$$

One-Sided p-value

`1 - pnorm(2.2)` ≈ 0.014

Two-Sided p-value

`2 * (1 - pnorm(2.2))` ≈ 0.028

Tests for Proportions: Two-Sample Example

From Pew Polling Data

53% of a random sample of 238 Democrats correctly identified Mitt Romney as Pro-Life versus 61% of 239 Republicans.

Sampling Model

Republicans: $X_1, \ldots, X_n \sim$ iid Bernoulli(p) independent of

Democrats: $Y_1, \ldots, Y_m \sim$ iid Bernoulli(q)

Sample Statistics

Sample Proportions: $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \widehat{q} = \frac{1}{m}\sum_{i=1}^{m} Y_i$

$\boxed{\text{Suppose I wanted to test } H_0\colon p = q}$

A More Efficient Estimator of the SE Under H_0

Don't Forget!

Standard Error (SE) means "std. dev. of sampling distribution" so you should know how to prove that:

$$SE(\widehat{p} - \widehat{q}) = \sqrt{\frac{p(1-p)}{n} + \frac{q(1-q)}{m}}$$

Under H_0: $p = q$

Don't know values of p and q: only that they are equal.

A More Efficient Estimator of the SE Under H_0

One Possible Estimate

$$\widehat{SE} = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n} + \frac{\widehat{q}(1-\widehat{q})}{m}}$$

A *Better* Estimate Under H_0

$$\widehat{SE}_{Pooled} = \sqrt{\widehat{\pi}(1-\widehat{\pi})\left(\frac{1}{n}+\frac{1}{m}\right)} \quad \text{where} \quad \widehat{\pi} = \frac{n\widehat{p}+m\widehat{q}}{n+m}$$

Why Pool?

If $p = q$, the two populations *are the same*. This means we can get a *more precise* estimate of the *common* population proportion by pooling. More data = Lower Variance $\implies$ better estimated SE.

Two-Sample Test for Proportions

Sampling Model

$X_1, \ldots, X_n \sim$ iid Bernoulli(p) indep. of $Y_1, \ldots, Y_m \sim$ iid Bernoulli(q)

Sample Statistics

Sample Proportions: $\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \widehat{q} = \frac{1}{m}\sum_{i=1}^{m} Y_i$

Null Hypothesis

$H_0\colon p = q \quad \Leftarrow \boxed{\text{ i.e. } p - q = 0}$

Pooled Estimator of SE under H_0

$$\widehat{\pi} = \frac{n\widehat{p} + m\widehat{q}}{n+m}, \quad \widehat{SE}_{Pooled} = \sqrt{\widehat{\pi}(1-\widehat{\pi})\,(1/n + 1/m)}$$

Test Statistic

$$T_n = \frac{\widehat{p} - \widehat{q}}{\widehat{SE}_{Pooled}} \approx N(0,1) \text{ under } H_0 \text{ provided } n \text{ and } m \text{ are large}$$

Two-Sample Example $H_0 : p = q$

53% of 238 Democrats knew Romney is Pro-Life vs. 61% of 239 Republicans

$$\widehat{\pi} = \frac{n\widehat{p} + m\widehat{q}}{n + m} = \frac{239 \times 0.61 + 238 \times 0.53}{239 + 238} \approx 0.57$$

$$\begin{aligned} \widehat{SE}_{Pooled} &= \sqrt{\widehat{\pi}(1 - \widehat{\pi})\,(1/n + 1/m)} = \sqrt{0.57 \times 0.43(1/239 + 1/238)} \\ &\approx 0.045 \end{aligned}$$

$$T_n = \frac{\widehat{p} - \widehat{q}}{\widehat{SE}_{Pooled}} = \frac{0.61 - 0.53}{0.045} \approx 1.78$$

One-Sided P-Value

`1 - pnorm(1.78)` ≈ 0.04

Two-Sided P-Value

`2 * (1 - pnorm(1.78))` ≈ 0.08