Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture #18 – Hypothesis Testing I

The Pepsi Challenge

Analogy between Hypothesis Testing and a Criminal Trial

Steps in a Hypothesis Test

The Pepsi Challenge

Our expert claims to be able to tell the difference between Coke and Pepsi. Let's put this to the test!

- Eight cups of soda
  - Four contain Coke
  - Four contain Pepsi
- The cups are randomly arranged
- How can we use this experiment to tell if our expert can *really* tell the difference?

The Results:

of Cokes Correctly Identified:

What do you think? Can our expert really tell the difference?

(a) Yes

(b) No

If you just guess randomly, what is the probability of identifying *all four cups of Coke correctly*?

- $\binom{8}{4} = 70$ ways to choose four of the eight cups.
- If guessing randomly, each of these is *equally likely*
- Only *one* of the 70 possibilities corresponds to correctly identifying all four cups of Coke.
- Thus, the probability is $1/70 \approx 0.014$

Probabilities if Guessing Randomly

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

If you're just guessing, what is the probability of identifying *at least* three Cokes correctly?

- Probabilities of mutually exclusive events sum.
- $P(\text{all four correct}) = 1/70$
- $P(\text{exactly 3 correct }) = 16/70$
- $P(\text{at least three correct}) = 17/70 \approx 0.24$

The Pepsi Challenge

- Even if you're just guessing randomly, the probability of correctly identifying three or more Cokes is around 24%
- In contrast, the probability of identifying *all four* Cokes correctly is only around 1.4% if you're guessing randomly.
- We should probably require the expert to get them all right...
- What if the expert gets them all wrong? This also has probability 1.4% if you're guessing randomly...

That was a hypothesis test! We'll go through the details in a moment, but first an analogy...

Criminal Trial

- The person on trial is either innocent or guilty (but not both!)
- "Innocent Until Proven Guilty"
- Only convict if evidence is "beyond a reasonable doubt"
- *Not Guilty* rather than Innocent
  - Acquit $\neq$ Innocent
- Two Kinds of Errors:
  - Convict the innocent
  - Acquit the guilty
- Convicting the innocent is a worse error. Want this to be rare even if it means acquitting the guilty.

Hypothesis Testing

- Either the null hypothesis H_0 or the alternative H_1 hypothesis is true.
- Assume H_0 to start
- Only reject H_0 in favor of H_1 if there is strong evidence.
- *Fail to reject* rather than Accept H_0
  - (Fail to reject H_0) $\neq$ (H_0 True)
- Two Kinds of Errors:
  - Reject true H_0 (Type I)
  - Don't reject false H_0 (Type II)
- Type I errors (reject true H_0) are worse: make them rare even if that means more Type II errors.

How is the Pepsi Challenge a Hypothesis Test?

Null Hypothesis H_0

Can't tell the difference between Coke and Pepsi: just guessing.

Alternative Hypothesis H_1

Able to tell which ones are Coke and which are Pepsi.

Type I Error – Reject H_0 even though it's true

Decide expert can tell the difference when she's really just guessing.

Type II Error – Fail to reject H_0 even though it's false

Decide expert just guessing when she really can tell the difference.

How do we carry out a hypothesis test?

Step 1 – Specify H_0 and H_1

- Pepsi Challenge: H_0 – our "expert" is guessing randomly
- Pepsi Challenge: H_1 – our "expert" can tell which is Coke

Step 2 – Choose a Test Statistic T_n

- T_n uses sample data to measure the plausibility of H_0 vs. H_1
- Pepsi Challenge: $T_n =$ Number of Cokes correctly identified
- Lots of Cokes correct $\Rightarrow$ implausible that you're just guessing

Step 3 – Calculate Distribution of T_n under H_0

- Under the null = Under H_0 = Assuming H_0 is true
- To carry out our test, need sampling dist. of T_n under H_0
- H_0 must be "specific enough" that we can do the calculation.
- Pepsi Challenge:

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

Step 4 – Choose a Critical Value c

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

- Pepsi Challenge: correctly identify many cokes $\Rightarrow$ implausible you're guessing at random.
- Decision Rule: reject H_0 if $T_n > c$, where c is the critical value.
- Choose c to ensure $P(\text{Type I Error})$ is small. But how small?
- Significance level α = max. prob. of Type I error we will allow
- Choose c so that if H_0 is true $P(T_n > c) \leq \alpha$
- Pepsi Challenge: if you are guessing randomly, then
  - $P(T_n > 3) = 1/70 \approx 0.014$
  - $P(T_n > 2) = 16/70 + 1/70 \approx 0.23$

How do we carry out a hypothesis test?

# Correct	0	1	2	3	4
Prob.	1/70	16/70	36/70	16/70	1/70

Step 1 – Specify Null Hypothesis H_0 and alternative Hypothesis H_1

Step 2 – Choose Test Statistic T_n

Step 3 – Calculate sampling dist of T_n under H_0

Step 4 – Choose Critical Value c

Step 5 – Look at the data: if $T_n > c$, reject H_0.

Pepsi Challenge

If $\alpha = 0.05$ we need $c = 3$ so that $P(T_n > 3) \leq \alpha$ under H_0.

Based on the results for our expert, would we reject H_0?

Lecture #19 – Hypothesis Testing II

Test for the mean of a normal population (variance known)

Relationship Between Confidence Intervals and Hypothesis Tests

P-values

One-Sided Tests

A Simple Example

Suppose $X_1, \ldots, X_{100} \sim$ iid $\text{N}(\mu, \sigma^2 = 9)$ and we want to test

$$H_0 \colon \mu = 2$$
$$H_1 \colon \mu \neq 2$$

Step 1 – Specify Null Hypothesis H_0 and alternative Hypothesis H_1 ✓

Step 2 – Choose Test Statistic T_n

If $\bar{X}$ is far from 2 then $\mu = 2$ is implausible. Why?

If $\bar{X}_n$ is far from 2, then $\mu = 2$ is implausible

Since $X_1, \ldots, X_{100} \sim$ iid N$(\mu, 9)$, if $\mu = 2$ then $\bar{X} \sim N(2, 0.09)$

$$
\begin{aligned}
P(a \leq \bar{X} \leq b) &= P\left(\frac{a-2}{3/10} \leq \frac{\bar{X}-2}{3/10} \leq \frac{b-2}{3/10}\right) \\
&= P\left(\frac{a-2}{0.3} \leq Z \leq \frac{b-2}{0.3}\right)
\end{aligned}
$$

where $Z \sim N(0,1)$ so we see that if $H_0\colon \mu = 2$ is true then

$$
\begin{aligned}
P(1.7 \leq \bar{X} \leq 2.3) &= P(-1 \leq Z \leq 1) \approx 0.68 \\
P(1.4 \leq \bar{X} \leq 2.6) &= P(-2 \leq Z \leq 2) \approx 0.95 \\
P(1.1 \leq \bar{X} \leq 2.9) &= P(-3 \leq Z \leq 3) > 0.99
\end{aligned}
$$

Step 2 – Choose Test Statistic T_n

- Reject $H_0\colon \mu = 2$ if the sample mean is far from 2.
- $\Rightarrow$ T_n should depend on the distance from $\bar{X}$ to 2, i.e. $|\bar{X} - 2|$.
- We can make our subsequent calculations much easier if we choose a scale for T_n that is convenient under H_0...

$$\mu = 2 \Rightarrow \quad \bar{X} - 2 \quad \sim \quad N(0, 0.09)$$

$$\frac{\bar{X} - 2}{0.3} \quad \sim \quad N(0, 1)$$

So we will set $T_n = \left| \dfrac{\bar{X} - 2}{0.3} \right|$

A Simple Example: $X_1, \ldots, X_{100} \sim \text{iid N}(\mu, \sigma^2 = 9)$

Step 1 – $H_0\colon \mu = 2,\ H_1\colon \mu \neq 2$ ✓

Step 2 – $T_n = \left|\dfrac{\bar{X} - 2}{0.3}\right|$ ✓

Step 3 – If $\mu = 2$ then $\left(\dfrac{\bar{X} - 2}{0.3}\right) \sim N(0, 1)$ ✓

Step 4 – Choose Critical Value c

(i) Specify significance level α.

(ii) Choose c so that $P(T_n > c) = \alpha$ under $H_0\colon \mu = 2$.

Choose c so that $P(T_n > c) = \alpha$ under H_0

$$T_n = \left|\frac{\bar{X}-2}{0.3}\right| \text{ and } \mu = 2 \implies \frac{\bar{X}-2}{0.3} \sim N(0,1)$$

$$\begin{aligned}
P\left(\left|\frac{\bar{X}-2}{0.3}\right| > c\right) &= \alpha \\
1 - P\left(\left|\frac{\bar{X}-2}{0.3}\right| \leq c\right) &= \alpha \\
P\left(\left|\frac{\bar{X}-2}{0.3}\right| \leq c\right) &= 1-\alpha \\
P\left(-c \leq \frac{\bar{X}-2}{0.3} \leq c\right) &= 1-\alpha
\end{aligned}$$

Hence: $c = \texttt{qnorm}(1-\alpha/2)$ which should look familiar!

A Simple Example: $X_1, \ldots, X_{100} \sim$ iid $\text{N}(\mu, \sigma^2 = 9)$

Step 1 – $H_0\colon \mu = 2,\ H_1\colon \mu \neq 2$ ✓

Step 2 – $T_n = \left|\dfrac{\bar{X} - 2}{0.3}\right|$ ✓

Step 3 – If $\mu = 2$ then $\left(\dfrac{\bar{X} - 2}{0.3}\right) \sim N(0,1)$ ✓

Step 4 – $c = \texttt{qnorm}(1 - \alpha/2)$ ✓

Step 5 – Look at the data: if $T_n > c$, reject H_0

- Suppose I choose $\alpha = 0.05$. Then $c \approx 2$.
- I observe a sample of 100 observations. Suppose $\bar{x} = 1.34$

$$T_n = \left|\frac{\bar{x} - 2}{0.3}\right| = \left|\frac{1.34 - 2}{0.3}\right| = 2.2$$

- Since $T_n > c$, I reject $H_0\colon \mu = 2$.

Reporting the Results of a Test

Our Example: $X_1, \ldots, X_{100} \sim$ iid $N(\mu, 1)$

- $H_0\colon \mu = 2$ vs. $H_1\colon \mu \neq 2$
- $T_n = |(\bar{X}_n - 2)/0.3|$
- $\alpha = 0.05 \implies c \approx 2$

Suppose $\bar{x} = 1.34$

Then $T_n = 2.2$. Since this is greater than c for $\alpha = 0.05$, we reject $H_0\colon \mu = 2$ at the 5% significance level.

Suppose instead that $\bar{x} = 1.82$

Then $T_n = 0.6$. Since this is less than c for $\alpha = 0.05$, we fail to reject $H_0\colon \mu = 2$ at the 5% signifcance level.

General Version of Preceding Example

$X_1, \ldots, X_n \sim$ iid $\mathrm{N}(\mu, \sigma^2)$ with σ^2 known and we want to test:

$$H_0\colon \mu = \mu_0$$
$$H_1\colon \mu \neq \mu_0$$

where μ_0 is some specified value for the population mean.

- $|\bar{X}_n - \mu_0|$ tells how far sample mean is from μ_0.
- Reject $H_0\colon \mu = \mu_0$ if sample mean is far from μ_0.
- Under $H_0\colon \mu = \mu_0$, $\dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$.
- Test statistic $T_n = \left|\dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right|$
- Reject $H_0\colon \mu = \mu_0$ if $T_n > \texttt{qnorm}(1 - \alpha/2)$

What is this test telling us to do?

Return to specific example where $H_0\colon \mu = 2$ vs. $H_1\colon \mu \neq 2$ and $X_1, \ldots, X_{100} \sim$ iid $\mathrm{N}(\mu, 1)$ with $\alpha = 0.05$:

$$\text{Reject } H_0 \quad \text{if} \quad \left|\frac{\bar{X}_n - 2}{0.3}\right| > 2$$

$$\text{Reject } H_0 \quad \text{if} \quad |\bar{X}_n - 2| > 0.6$$

$$\text{Reject } H_0 \quad \text{if} \quad (\bar{X}_n < 1.4) \text{ or } (\bar{X}_n > 2.6)$$

Reject $H_0\colon \mu = 2$ if $\bar{X}_n$ is far from 2. How far? Depends on choice of α along with sample size and population variance.

This looks suspiciously similar to a confidence interval...

$$\boxed{X_1, \ldots, X_n \sim \text{iid N}(\mu, \sigma^2) \text{ where } \sigma^2 \text{ is known}}$$

$$\boxed{T_n = \left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right|, \ c = \texttt{qnorm}(1-\alpha/2), \ \text{Reject } H_0\colon \mu = \mu_0 \text{ if } T_n > c}$$

Another way of saying this is don't reject H_0 if:

$$\begin{aligned}
(T_n \leq c) \quad &\iff \quad \left(\left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| \leq c\right) \iff \left(-c \leq \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \leq c\right) \\
&\iff \quad \left(\bar{X}_n - c \times \frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{X}_n + c \times \frac{\sigma}{\sqrt{n}}\right)
\end{aligned}$$

In other words, don't reject $H_0\colon \mu = \mu_0$ at significance level α if μ_0 lies inside the $100 \times (1-\alpha)\%$ confidence interval for μ.

CIs and Hypothesis Tests are Intimately Related

Our Simple Example

$X_1, \ldots, X_{100} \sim$ iid $\text{N}(\mu, \sigma^2 = 9)$ and observe $\bar{x} = 1.34$

Test $H_0\colon \mu = 2$ vs. $H_1\colon \mu \neq 2$ with $\alpha = 0.05$

$T_n = 2.2$, $c = \texttt{qnorm}(1 - 0.05/2) \approx 2$. Since $T_n > c$ we reject.

95% Confidence Interval for μ

$1.34 \pm 2 \times 3/10$ i.e. 1.34 ± 0.6 or equivalently $(0.74, 1.94)$

Another way to carry out the test…

Since 2 lies outside the 95% confidence interval for μ, if our significance level is $\alpha = 0.05$ we reject $H_0\colon \mu = 2$.

$X_1, \ldots X_{100} \sim$ iid $N(\mu_X, 1)$ and $Y_1, \ldots, Y_{100} \sim$ iid $N(\mu_Y, 1)$

Two researchers: $H_0\colon \mu = 2$ vs. $H_1\colon \mu \neq 2$ with $\alpha = 0.05$

Researcher 1

- $\bar{x} = 1.34$
- $T_n = 2.2 > 2$
- Reject $H_0\colon \mu_X = 2$

Researcher 2

- $\bar{y} = 11.3$
- $T_n = 31 > 2$
- Reject $H_0\colon \mu_Y = 2$

Both researchers would report "reject H_0 at the 5% level" but Researcher 2 found much stronger evidence against H_0...

What if we had chosen a different significance level α?

$$\boxed{T_n = 2.2, \quad c = \texttt{qnorm}(1 - \alpha/2), \quad \text{Reject } H_0 \colon \mu = 2 \text{ if } T_n > c}$$

$\alpha = 0.32 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.32/2) \approx 0.99$ Reject

$\alpha = 0.10 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.10/2) \approx 1.64$ Reject

$\alpha = 0.05 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.05/2) \approx 1.96$ Reject

$\alpha = 0.04 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.04/2) \approx 2.05$ Reject

$\alpha = 0.03 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.03/2) \approx 2.17$ Reject

$\alpha = 0.02 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.02/2) \approx 2.33$ Fail to Reject

$\alpha = 0.01 \quad \Rightarrow \quad c = \texttt{qnorm}(1 - 0.01/2) \approx 2.58$ Fail to Reject

Result of Test Depends on Choice of α!

$\alpha = 0.32$	$\Rightarrow$	Reject
$\alpha = 0.10$	$\Rightarrow$	Reject
$\alpha = 0.05$	$\Rightarrow$	Reject
$\alpha = 0.04$	$\Rightarrow$	Reject
$\alpha = 0.03$	$\Rightarrow$	Reject
$\alpha = 0.02$	$\Rightarrow$	Fail to Reject
$\alpha = 0.01$	$\Rightarrow$	Fail to Reject

- If you reject H_0 at a given choice of α, you would also have rejected at any larger choice of α.
- If you fail to reject H_0 at a given choice of α, you would also have failed to reject at any smaller choice of α.

Question

If α is large enough we will reject; if α is small enough, we won't. Where is the dividing line between reject and fail to reject?

P-Value: Dividing Line Between Reject and Fail to Reject

$$\boxed{T_n = 2.2, \quad c = \texttt{qnorm}(1 - \alpha/2), \quad \text{Reject } H_0\colon \mu = 2 \text{ if } T_n > c}$$

Question

Given that we observed a test statistic of 2.2, what choice of α would put us just at the cusp of rejecting H_0?

Answer

Whichever α makes $c = 2.2$! At this α we just barely fail to reject.

Calculating the P-value

Definition of a P-value

Significance level α such that the critical value c exactly equals the observed value of the test statistic. Equivalently: α that lies exactly on boundary between Reject and Fail to Reject.

Our Example

The observed value of the test statistic is 2.2 and the critical value is $\texttt{qnorm}(1-\alpha/2)$, so we need to solve:

$$
\begin{aligned}
2.2 &= \texttt{qnorm}(1-\alpha/2) \\
\texttt{pnorm}(2.2) &= \texttt{pnorm}\left(\texttt{qnorm}\left(1-\alpha/2\right)\right) \\
\texttt{pnorm}(2.2) &= 1-\alpha/2 \\
\alpha &= 2\times[1-\texttt{pnorm}(2.2)] \approx 0.028
\end{aligned}
$$

How to use a p-value?

Alternative to Steps 4–5

Rather than choosing α, computing critical value c and reporting "Reject" or "Fail to Reject" at $100 \times \alpha\%$ level, just report p-value.

Example From Previous Slide

P-value for our test of $H_0 \colon \mu = 2$ against $H_1 \colon \mu \neq 2$ was ≈ 0.028

Using P-value to Test H_0

Using the p-value we can test H_0 for any α without doing any new calculations! For p-value $< \alpha$ reject; for p-value $\geq \alpha$ fail to reject.

Strength of Evidence Against H_0

P-value measures strength of evidence against the null. Smaller p-value $=$ stronger evidence against H_0. P-value does not measure size of effect.

One-sided Test: Different Decision Rule

Same Example as Above

$X_1, \ldots, X_{100} \sim$ iid $\text{N}(\mu, 1)$ and $H_0\colon \mu = 2$.

Three possible alternatives:

Two-sided	One-sided ($<$)	One-sided ($>$)
$H_1\colon \mu \neq 2$	$H_1\colon \mu < 2$	$H_1\colon \mu > 2$

Three corresponding decision rules:

- Two-sided: reject $\mu = 2$ whenever $|\bar{X}_n - 2|$ is too large.
- One-sided ($<$): only reject $\mu = 2$ if $\bar{X}_n$ is far below 2.
- One-sided ($>$): only reject $\mu = 2$ if $\bar{X}_n$ is far above 2.

One-sided ($>$) Example: $X_1, \ldots, X_{100} \sim$ iid N$(\mu, 1)$

Null and Alternative

Test $H_0\colon \mu = 2$ against $H_0\colon \mu > 2$ with $\alpha = 0.05$.

Test Statistic

Drop absolute value for one-sided test: $T_n = \dfrac{\bar{X}_n - 2}{0.3}$

Decision Rule

Reject $H_0\colon \mu = 2$ if test statistic is large and positive: $T_n > c$

Critical Value

Choose c so that $P(\text{type I error}) = P(T_n > c|\mu = 2) = 0.05$

Under H_0, $T_n \sim N(0,1)$

If $Z \sim N(0,1)$ what value of c ensures $P(Z > c) = 0.05$?

One-sided ($<$) Example: $X_1, \ldots, X_{100} \sim$ iid N$(\mu, 1)$

Null and Alternative

Test $H_0\colon \mu = 2$ against $H_0\colon \mu < 2$ with $\alpha = 0.05$.

Test Statistic

Drop absolute value for one-sided test: $T_n = \dfrac{\bar{X}_n - 2}{0.3}$

Decision Rule

Reject $H_0\colon \mu = 2$ if test statistic is large and negative: $T_n < c$

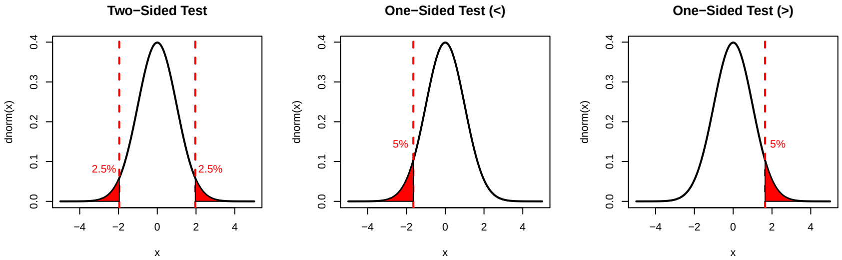
Critical Value

Choose c so that $P(\text{type I error}) = P(T_n < c|\mu = 2) = 0.05$

Under H_0, $T_n \sim N(0,1)$

If $Z \sim N(0,1)$ what value of c ensures $P(Z < c) = 0.05$?

Critical Values – Two-sided vs. One-sided Tests: $\alpha = 0.05$



Two-Sided

Splits $\alpha = 0.05$ between two tails: $c = \texttt{qnorm}(1 - 0.05/2) \approx 1.96$

One-Sided

One tail: $c = \texttt{qnorm}(0.05) \approx -1.64$ for $(<)$; $\texttt{qnorm}(0.95) \approx 1.64$ for $(>)$

Example: $X_1, \ldots, X_{100} \sim$ iid $\mathrm{N}(\mu, 1), \alpha = 0.05$

$$\boxed{\text{Suppose } \bar{x} = 1.5 \implies (\bar{x} - 2)/0.3 \approx -1.67}$$

Two-sided

$H_1: \mu \neq 2$

Reject if $T_n > 1.96$

$T_n = 1.67$

Fail to reject

One-sided ($<$)

$H_1: \mu < 2$

Reject if $T_n < -1.64$

$T_n = -1.67$

Reject

One-sided ($>$)

$H_1: \mu > 2$

Reject if $T_n > 1.64$

$T_n = -1.67$

Fail to reject

- If One-sided ($<$) rejects, then one-sided ($>$) doesn't and vice-versa.
- Two-sided and one-sided sometimes agree but sometimes disagree.
- One-sided test is "less stringent."

Testing $H_0\colon \mu = \mu_0$ when $X_1, \ldots, X_n \sim$ iid $\text{N}(\mu, \sigma^2)$

Two-Sided

Reject H_0 whenever $\left|\dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| > \texttt{qnorm}(1-\alpha/2)$

One-Sided ($<$)

Reject H_0 whenever $\dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} < \texttt{qnorm}(\alpha)$

One-Sided ($>$)

Reject H_0 whenever $\dfrac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > \texttt{qnorm}(1-\alpha)$

One-sided P-value

- Only makes sense to calculate one-sided p-value when sign of test stat. agrees with alternative:
  - Preceding example: $T_n = -1.67$
  - Calculate p-value for test vs. $H_1 : \mu < 2$ but not $H_1 : \mu > 2$
- Just as in two-sided test, p-value equals value of α for which c exactly equals the observed test statistic:
  - $c = \texttt{qnorm}(\alpha)$ for $(<)$
  - $c = \texttt{qnorm}(1 - \alpha)$ for $(>)$
  - Example: $-1.67 = \texttt{qnorm}(\alpha) \iff \alpha = 0.047$
- Use and report one-sided p-value in same way as two-sided p-value

Final Notes on One-sided vs. Two-sided Tests

- Two-sided test is the default.
- Don't use one-sided unless you have a good reason!
- Relationship between CI and test only holds for two-sided.
- Why and when should we consider a one-sided test?
  - Suppose we know *a priori* that $\mu < 2$ is crazy/uninteresting
  - Test of $H_0: \mu = 2$ against $H_1: \mu > 2$ with significance level α has lower type II error rate than test against $H_1: \mu \neq 2$.
- If you use a one-sided test you must choose ($>$) or ($<$) *before looking at the data*. Otherwise the results are invalid.

Roadmap

Next Time

More examples of hypothesis testing using relationship to CIs to help us avoid re-inventing the wheel.

Building Intuition

Now that you know a simple example of a hypothesis test and its relationship to a CI, think about the following:

- If we reject H_0 does that mean that H_0 is false?
- How does testing relate to random sampling?
- How does critical value of two-sided test relate to width of CI?
- In a given test, which is larger: the one-sided or two-sided p-value?