Economics 103 – Statistics for Economists

Minsu Chang

University of Pennsylvania

Lecture 20

Hypothesis Testing II

What if the null is false?

Alternative hypothesis: H_1

The *negation* of the null hypothesis.

Examples:

1.
   - H_0: This parameter equals 5.
   - H_1: This parameter does *not* equal 5.
2.
   - H_0: There is no difference between these two groups.
   - H_1: There *is* a difference between these two groups.

Sometimes we only care about *certain kinds* of violations of H_0...

One-sided vs. Two-sided Alternative

Let θ be a population parameter and θ_0 be a specified constant.

Null Hypothesis

- $H_0 : \theta = \theta_0$

Two-sided Alternative

- $H_1 : \theta \neq \theta_0$

One-sided Alternative

Two possibilities, depending on the problem at hand:

- $H_1 : \theta > \theta_0$
- $H_1 : \theta < \theta_0$

Example: Suing McDonald's

A class action lawsuit claims that McDonald's has been understating the caloric content of the "Big Mac," misleading consumers into thinking the sandwich is healthier than it really is. McDonald's claims the sandwich contains 550 kcal on average.

Suppose you're the judge in this case. What is your alternative hypothesis?

(a) $H_1: \mu \neq 550$ kcal

(b) $H_1: \mu < 550$ kcal

(c) $H_1: \mu > 550$ kcal

(d) $H_1: \mu = 550$ kcal

Example: Quality Control at McDonald's

You are a senior manager at McDonald's and are concerned that franchises may be deviating from company policy on the calorie count of a Big Mac sandwich, which is supposed to be 550 kcal on average. Because intervening is costly, you will only take action is there is strong evidence of deviation from company policy.

What is your alternative hypothesis?

(a) $H_1 : \mu \neq 550$ kcal

(b) $H_1 : \mu < 550$ kcal

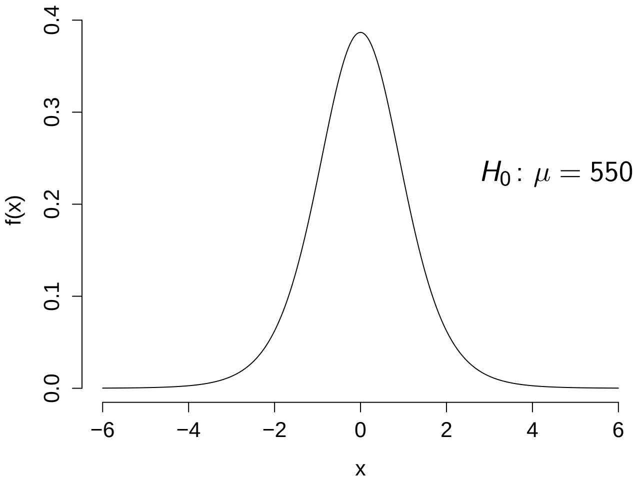
(c) $H_1 : \mu > 550$ kcal

(d) $H_1 : \mu = 550$ kcal

Decision Rule: When should we reject H_0?

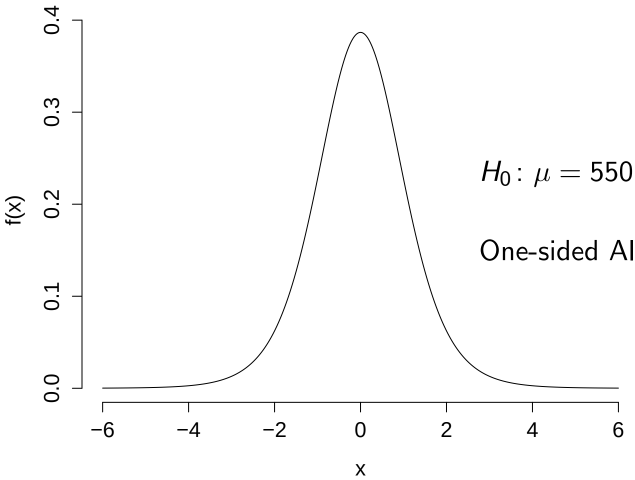
- Test statistic: RV with known sampling distribution under H_0
- McDonald's Example: $T_n = 3(\bar{X} - 550)/S$
- *Random* since $\bar{X}$ and S are RVs under random sampling: functions of $X_1, \ldots, X_9$.
- Observed dataset: *realizations* $x_1, \ldots, x_9$ of RVs $X_1, \ldots, X_9$
- Plug in observed data to get estimates (constants) $\bar{x}$ and s.
- Plug these into the formula for the test statistic to get a *number* – this is a *realization* of T_n
- Depending on this number, decide whether to reject H_0.

What Form Should the Decision Rule Take?



$$H_0: \mu = 550 \Rightarrow \frac{\bar{X} - 550}{S/3} \sim t(8)$$

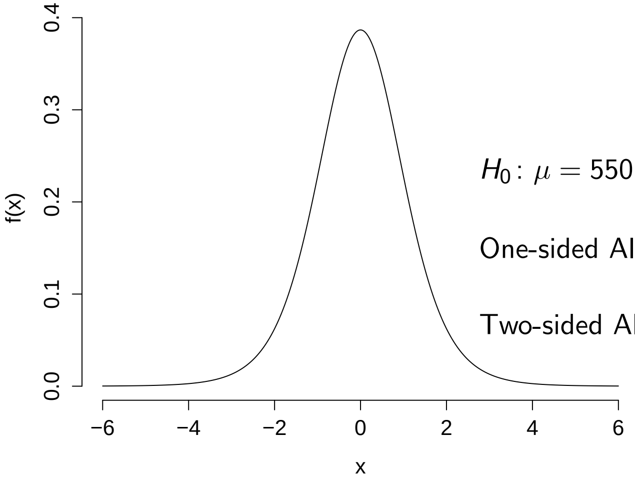
What Form Should the Decision Rule Take?



$$H_0: \mu = 550 \Rightarrow \frac{\bar{X} - 550}{S/3} \sim t(8)$$

One-sided Alternative $H_1: \mu > 550$

What Form Should the Decision Rule Take?



$$H_0\colon \mu = 550 \Rightarrow \frac{\bar{X} - 550}{S/3} \sim t(8)$$

One-sided Alternative $H_1\colon \mu > 550$

Two-sided Alternative $H_1\colon \mu \neq 550$

Example: Suing McDonald's

The plaintiffs allege that McDonald's has *understated* the true caloric content of a Big Mac: it's actually *greater* than 550 kcal. Suppose the plaintiffs are right. Then what sort of value should we expect the test statistic $3(\bar{X} - 550)/S$ to take on?

(a) A value *less* than zero.

(b) A value close to zero.

(c) A value *greater* than zero.

Example: Quality Control at McDonald's

The senior manager is worried that franchises are deviating from company policy that Big Macs should contain approximately 550 kcal. If the franchises *are* deviating, what sort of of value should we expect the test statistic $3(\bar{X} - 550)/S$ to take on?

(a) A value *less* than zero.

(b) A value close to zero.

(c) A value *greater* than zero.

(d) A value different from zero but we can't tell whether it will be positive or negative.

What Form Should the Decision Rule Take?

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

Common Null Hypothesis $H_0\colon \mu = 550$

Under H_0, $T_n = \sqrt{n}(\bar{X}_n - 550)/S \sim t(n-1)$

One-sided Alternative $H_1\colon \mu > 550$

Reject H_0 if T_n is "too big"

Two-sided Alternative $H_1\colon \mu \neq 550$

Reject H_0 if T_n is "too big" or "too small"

But how big of a discrepancy is "big enough" to reject?

Two Kinds of Mistakes in Hypothesis Testing

Type I Error

- Rejecting the null when it's actually true.
- $P(\text{Type I Error}) = \alpha$ $\boxed{\alpha = \text{“Significance Level” of Test}}$

Type II Error

- Failing to reject the null when it's false.
- $P(\text{Type II Error}) = \beta$ $\boxed{1 - \beta = \text{“Power” of Test}}$

Important!

Hypothesis testing *controls* probability of a Type I error since this is assumed to be the *worse* kind of mistake: convicting the innocent.

Construct a Decision Rule to *Fix* α at User-Chosen Level

Critical Value c_α

- Threshold for rejecting H_0
- Chosen so that $P(\text{Reject } H_0 | H_0 \text{ is True}) = \alpha$
- Depends on *both* α *and* the alternative hypothesis.

One-Sided Alternative

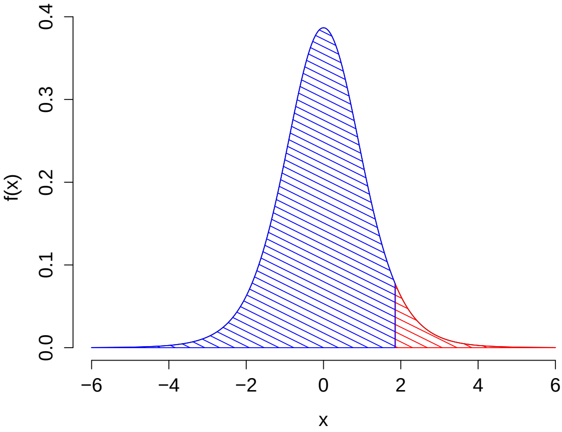
Reject H_0 if $T_n >$ Critical Value

Two-Sided Alternative

Reject H_0 if $|T_n| >$ Critical Value

Example: One-sided Alternative $H_1: \mu > 550$

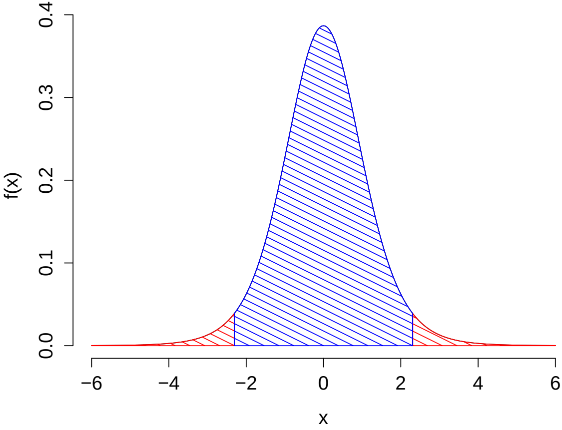
The critical value is chosen to reflect both the alternative hypothesis and the significance level.



One-sided Critical Value: `qt(`$1-\alpha$`, df = `$n-1$`)`

Example: Two-sided Alternative $H_1: \mu \neq 550$

The critical value is chosen to reflect both the alternative hypothesis and the significance level.

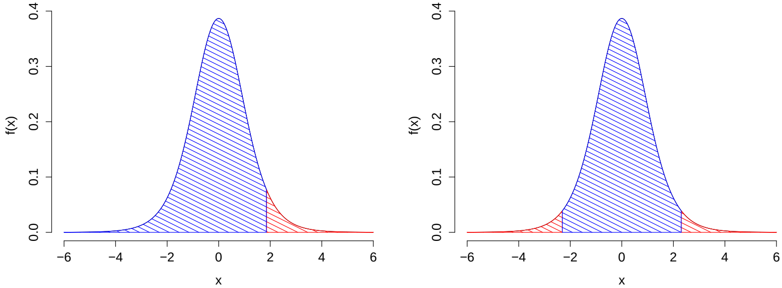


Two-sided Critical Value: `qt(`$1-\alpha/2$`, df = `$n-1$`)`

Suppose, for example, $\alpha = 0.05$, $n = 9$

`qt(0.95, df = 8)` ≈ 1.86

`qt(0.975, df = 8)` ≈ 2.3



One-sided Alternative: Reject H_0 if $3(\bar{X}_n - 550)/S \geq 1.86$

Two-sided Alternative: Reject H_0 if $\left|3(\bar{X}_n - 550)/S\right| \geq 2.3$

McDonald's Example

Suppose $n = 9$, $\bar{x} = 563$, $s = 34$. What is the value of our test statistic?

McDonald's Example

Suppose $n = 9$, $\bar{x} = 563$, $s = 34$. What is the value of our test statistic?

$$\frac{563 - 550}{34/\sqrt{9}} = \frac{13}{34/3} \approx 1.14$$

McDonald's Example: $\alpha = 0.05$

Recall that:

$$\texttt{qt(0.95, df = 8)} \approx 1.86$$

$$\texttt{qt(0.975, df = 8)} \approx 2.3$$

Based on an observed test statistic of 1.14, would we reject H_0 against the one-sided alternative at the 5% significance level?

(a) Yes

(b) No

(c) Not Sure

McDonald's Example: $\alpha = 0.05$

Recall that:

$$\texttt{qt(0.95, df = 8)} \approx 1.86$$

$$\texttt{qt(0.975, df = 8)} \approx 2.3$$

Based on an observed test statistic of 1.14, would we reject H_0 against the two-sided alternative at the 5% significance level?

(a) Yes

(b) No

(c) Not Sure

Reporting the Results of a Hypothesis Test

Lawsuit Example

The judge *failed to reject* the null hypothesis that $\mu = 550$ against the one-sided alternative $\mu > 550$ at the 5% significance level.

Quality Control Example

The senior manager *failed to reject* the null hypothesis that $\mu = 550$ against the two-sided alternative at the 5% significance level.

Interpretation

In each of these two cases, there was *insufficient evidence* against the initial assumption $\mu = 550$ *given the significance level used*.

But what if we have used a *different* significance level?

The P-Value of a Hypothesis Test

Two Equivalent Definitions:

1. Given the value we calculated for our test statistic, what is the *smallest* α at which we would have rejected the null?
2. Under the null, what is the probability of observing a test statistic *at least as extreme* as the one we *actually* observed?

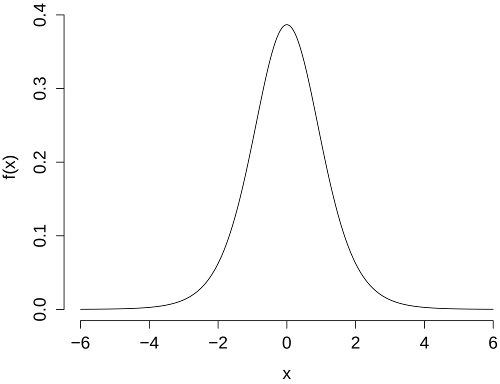
Why Report P-Values?

- More informative than reporting α and Reject/Fail to Reject
- E.g. a p-value of 0.03 means we would have rejected the null for any $\alpha \geq 0.03$ and failed to reject it for any $\alpha < 0.03$

P-Value Depends on Which Alternative We Have Specified!

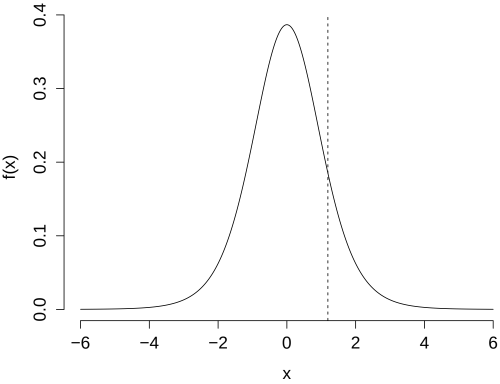
What is the p-value? (One-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the one-sided p-value?



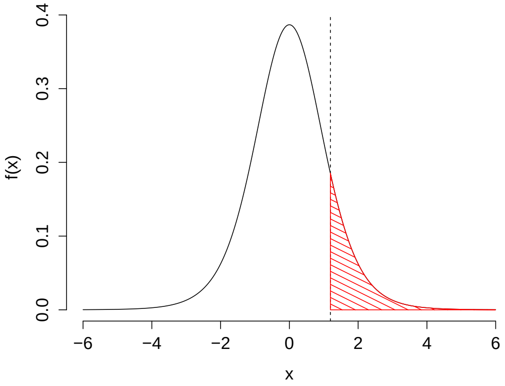
What is the p-value? (One-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the one-sided p-value?



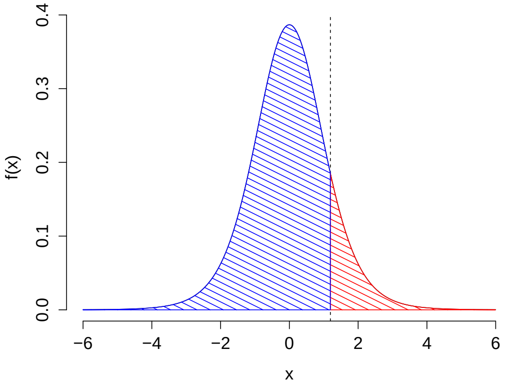
What is the p-value? (One-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the one-sided p-value?



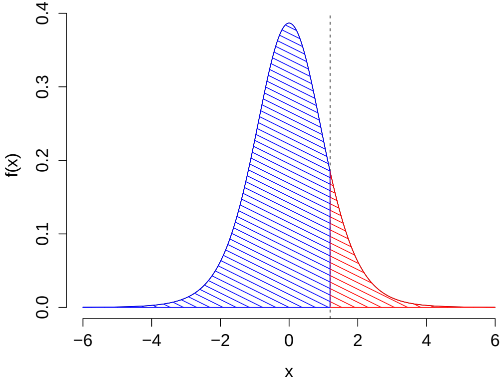
What is the p-value? (One-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the one-sided p-value?



What is the p-value? (One-sided Test)

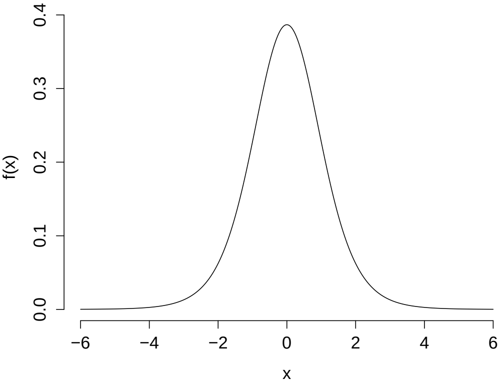
Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the one-sided p-value?



`1 - pt(1.14, df = 8)` ≈ 0.14

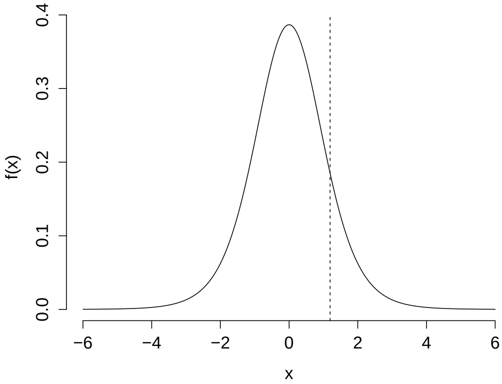
What is the p-value? (Two-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



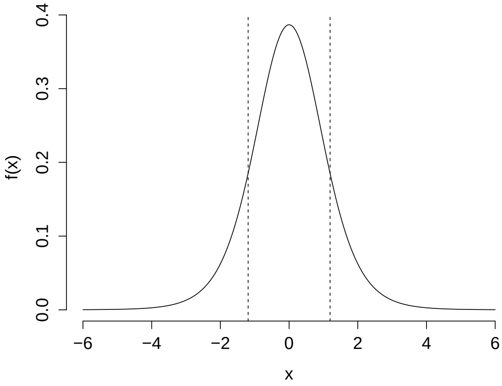
What is the p-value? (Two-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



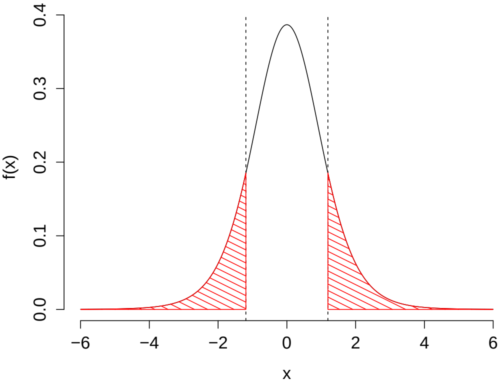
What is the p-value? (Two-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



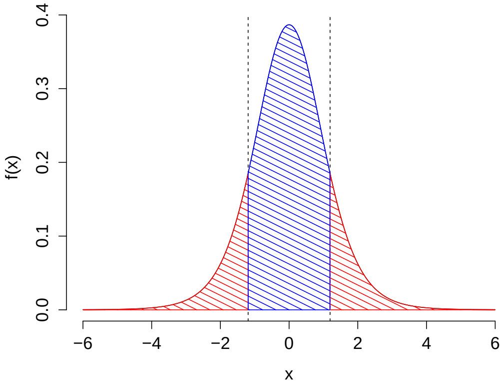
What is the p-value? (Two-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



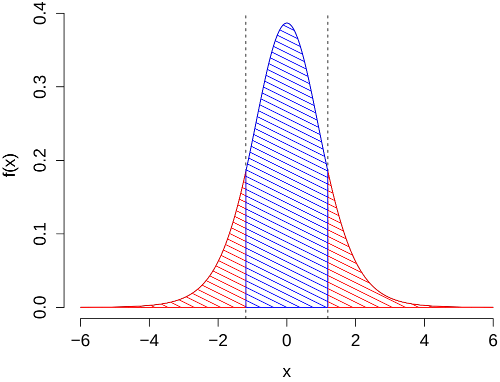
What is the p-value? (Two-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



What is the p-value? (Two-sided Test)

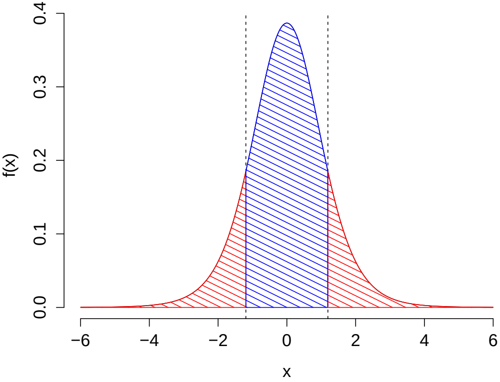
Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



`2 * pt(-1.14, df = 8)` ≈ 0.28

What is the p-value? (Two-sided Test)

Recall: p-value is *smallest significance level* at which our observed test statistic would cause us to reject H_0. Test statistic is 1.14. What is the two-sided p-value?



`2 * pt(-1.14, df = 8)`≈ 0.28 This is twice the one-sided p-value!

Two-sided Test is More Stringent

P-value measures strength of evidence against H_0

Lower p-value means stronger evidence.

(Two-sided p-value) $= 2 \times$ (one-sided p-value)

Reject H_0 based on two-sided test $\implies$ Reject H_0 based on appropriate one-sided test. The converse is *false*.

Steps in Hypothesis Testing

1. Specify Null and Alternative Hypotheses
2. Identify a Test Statistic: a function of the data that has a known sampling distribution under the null.
3. Specify a Decision Rule and a Critical Value so the Type I Error Rate equals α.

Alternative to Step 3

Calculate P-Value: the minimum significance level (α) at which we would reject H_0 given the observed data.

How to Handle Other Examples?

You already know lots of sampling distributions! Testing is very similar to constructing confidence intervals in that the steps are always the same, and the only thing that differs is *which* sampling distribution we work with. We'll look at more examples next time.