Review Session 8

API-201, 11.12.21
Sophie Hill

Agenda

- 3 methods of hypothesis testing
- Guidance on using weights for the final exercise

Example

Suppose we want to test if the **average height** of students in school A and school B is the same.

We take a random sample from each school and find that the students in school B are **6cm taller** on average.

Is this difference statistically significant at alpha = 0.05?

Example

Suppose we want to test if the **average height** of students in school A and school B is the same.

We take a random sample from each school and find that the students in school B are **6cm taller** on average.

Is this difference statistically significant at alpha = 0.05?

	School A	**School B**
Sample std dev	12.2	17.9
Sample size	28	30

What is the standard error of the difference in means?

Example

Suppose we want to test if the **average height** of students in school A and school B is the same.

We take a random sample from each school and find that the students in school B are **6cm taller** on average.

Is this difference statistically significant at alpha = 0.05?

	School A	**School B**
Sample std dev	12.2	17.9
Sample size	28	30

What is the standard error of the difference in means?

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} =$$

Example

Suppose we want to test if the **average height** of students in school A and school B is the same.

We take a random sample from each school and find that the students in school B are **6cm taller** on average.

Is this difference statistically significant at alpha = 0.05?

	School A	**School B**
Sample std dev	12.2	17.9
Sample size	28	30

What is the standard error of the difference in means?

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{12.2^2}{28} + \frac{17.9^2}{30}} = 4$$

Example

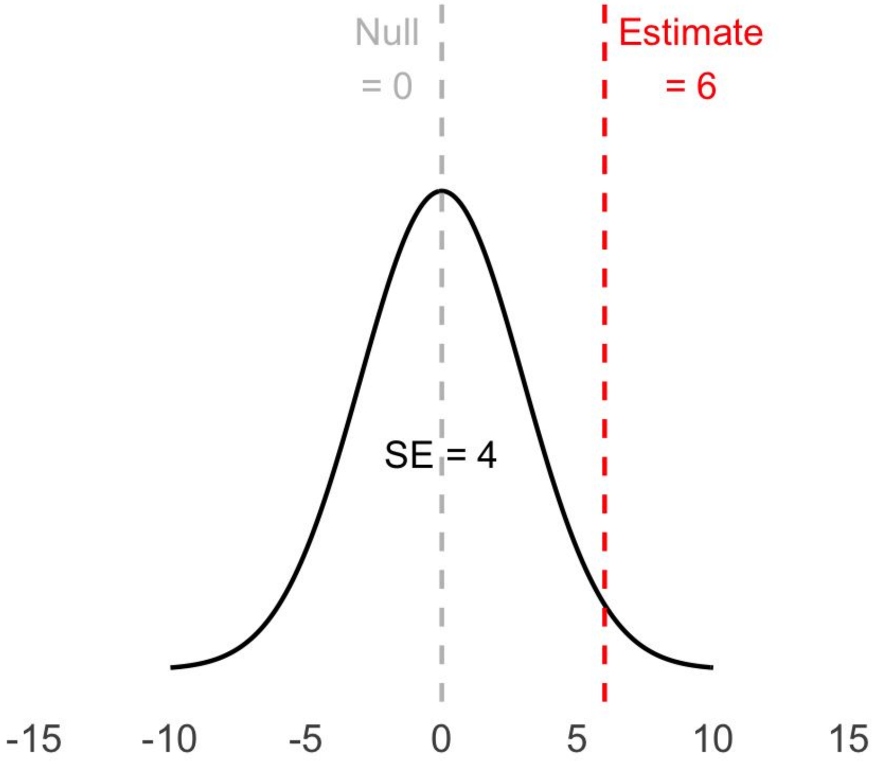
Null hypothesis

$H_o: \mu_B - \mu_A = 0$

Estimate

$\bar{x}_B - \bar{x}_A = 6cm$

SE = 4cm



Method 1: p-value

T-statistic

p-value

Method 1: p-value

T-statistic

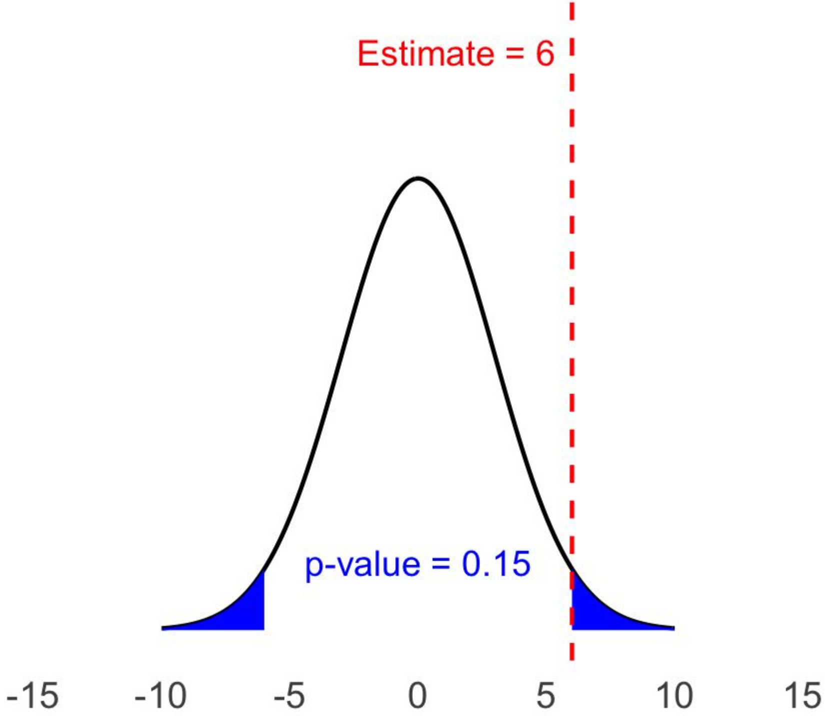
= (est - null)/SE

= (6 - 0)/4 = **1.5**

p-value
= =TDIST(ABS(1.5), 28-1, 2)
= 0.15

$p = 0.15 > 0.05 = \alpha$

→ **Fail to reject**



Method 2: rejection region

α = 0.05, so usually our multiplier would be 1.96

Since we are using the t-distribution, it's going to be a bit bigger.

You calculate this in Excel with =T.INV.2T(0.05, 27) = 2.05

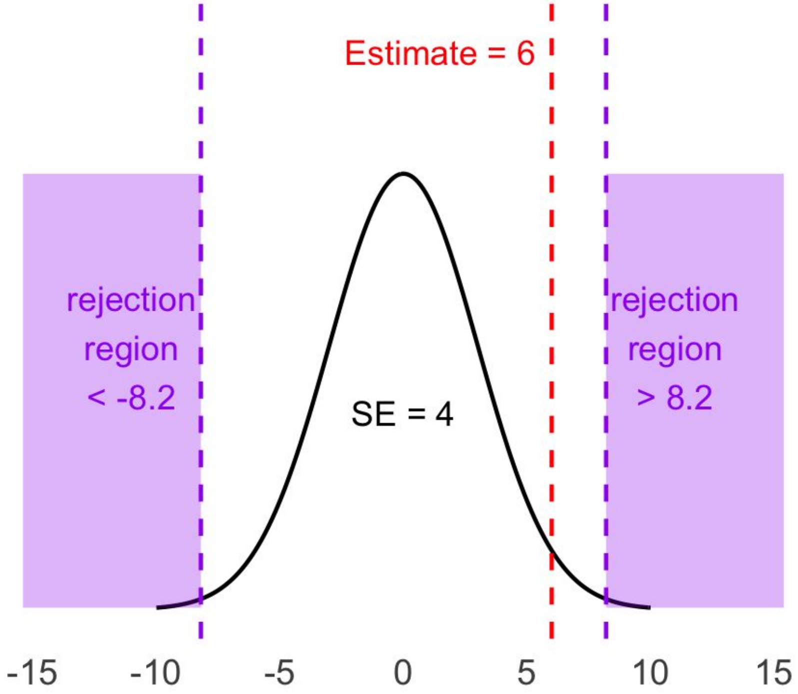
Method 2: rejection region

The rejection region is consists of values *more extreme* than:

Null value ± 2.05*SE
= 0 ± 2.05*4
= ± 8.2

Our estimate is 6cm, which does not fall in the rejection region.

→ **Fail to reject**



Method 3: confidence interval

We use the same multiplier as before: 2.05

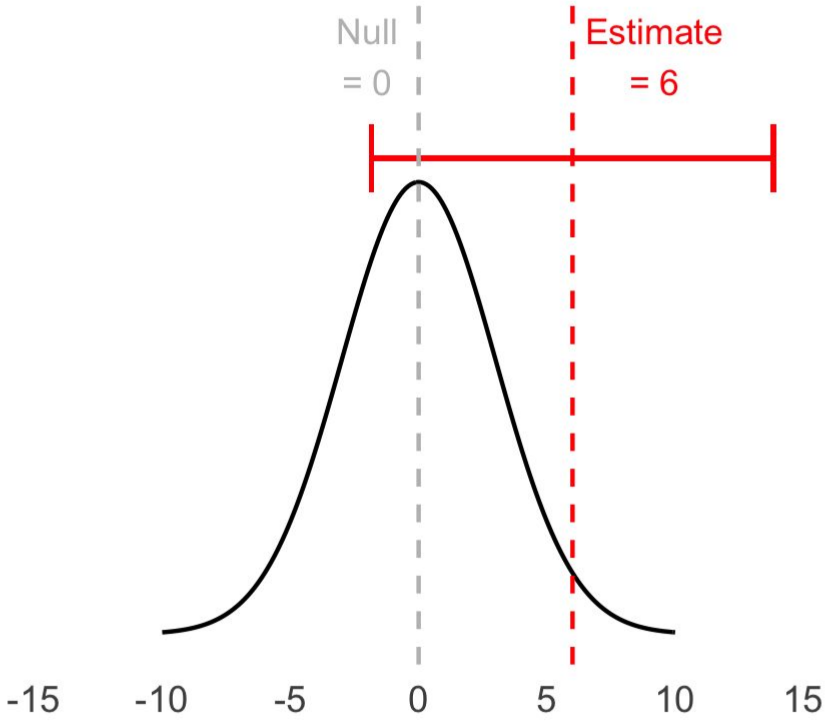
But this time we are putting the estimate in the middle, not the null value!

The confidence interval is defined by:

Estimate ± 2.05*SE
= 6 ± 2.05*4
= [-2.2, 14.2]

The confidence interval includes our null value of 0cm.

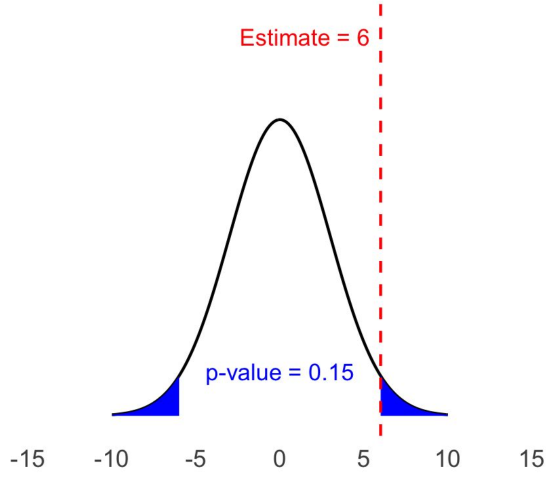
Fail to reject



Three methods

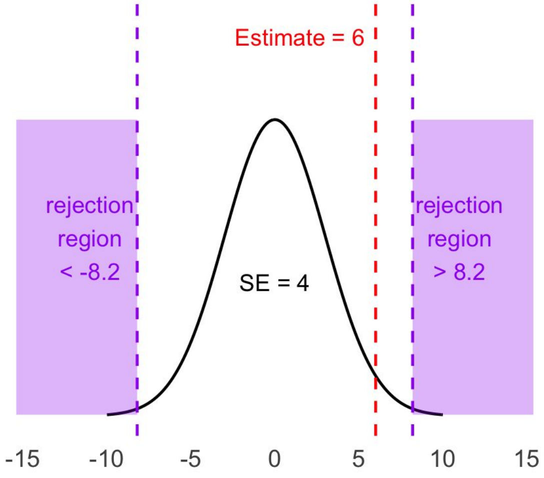
p-value

Is the p-value of my estimate < alpha ?



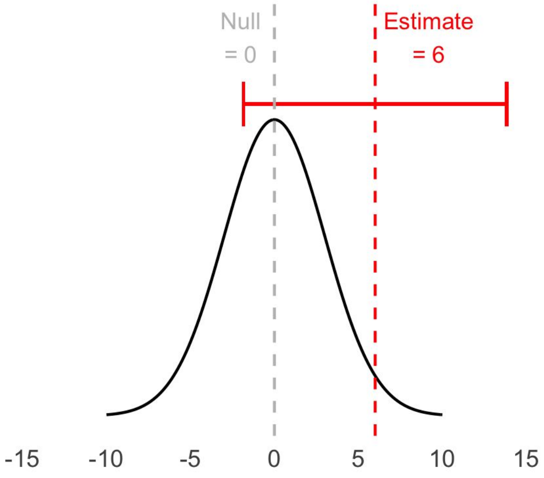
Rejection region

Does my estimate fall in the rejection region?



Confidence interval

Does my confidence interval exclude the null?



Practice problem

Here are the results of poll evaluating support for drilling for oil and natural gas off the coast of California.

(a) What proportion of college grads and non-college grads support drilling for oil and natural gas?

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

OIS, Exercise 6.25

Practice problem

Here are the results of poll evaluating support for drilling for oil and natural gas off the coast of California.

(a) What proportion of college grads and non-college grads support drilling for oil and natural gas?

q_c-hat = 154/438 = 0.352

q_{nc}-hat = 132/389 = 0.339

	College Grad Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

OIS, Exercise 6.25

Practice problem

Here are the results of poll evaluating support for drilling for oil and natural gas off the coast of California.

(b) Is the difference between these proportions statistically significant at the 5% level?

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

OIS, Exercise 6.25

Practice problem

Here are the results of poll evaluating support for drilling for oil and natural gas off the coast of California.

(b) Is the difference between these proportions statistically significant at the 5% level?

q_c-hat = 0.352 q_{nc}-hat = 0.339

H_0: $q_c - q_{nc} = 0$

Estimate = q_c-hat - q_{nc}-hat =

0.352-0.339 = 0.013

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

OIS, Exercise 6.25

Practice problem

Here are the results of poll evaluating support for drilling for oil and natural gas off the coast of California.

(b) Is the difference between these proportions statistically significant at the 5% level?

$$\sqrt{\frac{q_1(1-q_1)}{n_1}+\frac{q_2(1-q_2)}{n_2}}$$
$$=\sqrt{\frac{0.352(1-0.352)}{438}+\frac{0.339(1-0.339)}{389}}$$
$$=0.0331$$

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

OIS, Exercise 6.25

Practice problem

Here are the results of poll evaluating support for drilling for oil and natural gas off the coast of California.

(b) Is the difference between these proportions statistically significant at the 5% level?

Z-score = (est-null)/SE

= (0.013 - 0)/0.0331

= 0.393

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

OIS, Exercise 6.25

Method 1: p-value

Alpha = 0.05 Estimate = 0.013 SE = 0.0331 Z = 0.393

P–value = ?

Method 1: p-value

Alpha = 0.05 Estimate = 0.013 SE = 0.0331 Z = 0.393

P-value = 2*(1-NORM.S.DIST(ABS(0.393), TRUE))

= 0.69

Conclusion?

Method 1: p-value

Alpha = 0.05 Estimate = 0.013 SE = 0.0331 Z = 0.393

P-value = 2*(1-NORM.S.DIST(ABS(0.393), TRUE))

= 0.69

Conclusion? Fail to reject!

Method 2: rejection region

Alpha = 0.05 Estimate = 0.013 SE = 0.0331 Z = 0.393

Rejection region is the values more extreme than:

Null value +/- 1.96*SE = 0 +/- 1.96*0.0331 = +/- 0.0649

Our estimate = 0.013, which is not bigger than 0.0649 or smaller than -0.0649

Fail to reject!

Method 3: confidence interval

Alpha = 0.05 Estimate = 0.013 SE = 0.0331 Z = 0.393

The confidence interval is given by:

Estimate +/- 1.96*SE = 0.013 +/- 1.96*0.0331 = 0.013 +/- 0.0649

= [-0.0519, 0.0779]

Our confidence interval contains the null value of 0.

Fail to reject!