Math 361

Cause and Effect Relationships

Inv. 3.2-3.4

Terminology for 2 variables

Explanatory variable: the variable we think might explain changes in the response variable

Response variable: the outcome of interest

Inv. 3.1: is the rate of hearing loss increasing over time?

Response variable = hearing loss (yes or no)

Explanatory variable = time period (1994 vs. 2006)

Identify the **response** and **explanatory** variables for the following research questions

- Were California residents more or less likely to have been born in California (i.e., native Californians) back in 1950 or in 2000?

- Does the size of the crowd (sold out or not) at a basketball game influence whether the home team wins or not?

- Does taking a fish oil supplement reduce blood pressure?

Solutions

- Were California residents more or less likely to have been born in California (i.e., native Californians) back in 1950 or in 2000?

RV = native Californian (yes or no)

EV = year (1950 or 200)

- Does the size of the crowd (sold out or not) at a basketball game influence whether the home team wins or not?

RV = home team won (yes or no)

EV = size of crowd (sold out or not)

- Does taking a fish oil supplement reduce blood pressure?

RV = low in blood pressure (yes or no)

EV = take fish oil (yes or no)

Inv. 3.2: Nightlights and Myopia

Is there an association between near-sightedness and the use of nightlights with infants?

Response variable?

Explanatory variable?

H_0:

H_a:

Inv. 3.2: two-way table

	Some light	Darkness
Near-sighted	188	18
Not near-sighted	119	154
Total	307	172

Inv. 3.2: two sample z-test

Rossman/Chance Applet Collection

Theory-Based Inference

Scenario: Two proportions

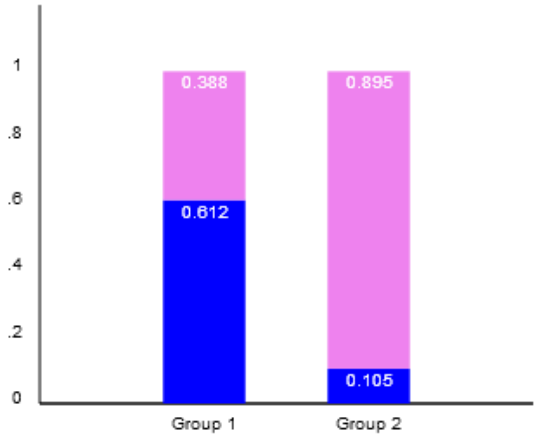
Paste Data

Group 1		Group 2	
n:	307	n:	172
count:	188	count:	18
sample $\hat{p}$:	0.612	sample $\hat{p}$:	0.104

Calculate

Reset

Sample Data



(Group1 - Group2)

$\hat{p}_1 - \hat{p}_2 = 0.508$

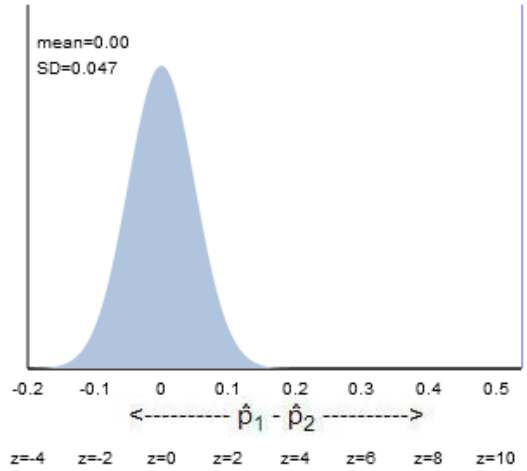
Theory-Based Inference

Test of significance

H_o: $\pi_1 - \pi_2$ = 0

H_a: $\pi_1 - \pi_2$ > 0

Calculate



standardized statistic z = 10.77

p-value 0.0000

Conclusion

From two-sample z-test, we got a p-value of about 0.

What do you conclude about nightlights and nearsightedness?

Inv. 3.2: part e

With a p-value of about 0, and assuming our samples are representative, we conclude that the proportion of near-sightedness increased when nightlights were used with infants.

Inv. 3.2: part e

With a p-value of about 0, and assuming our samples are representative, we conclude that the proportion of near-sightedness increased when nightlights were used with infants.

Would you say nightlight use with infants ***caused*** *the increase in near-sightedness?*

Inv. 3.2: part e

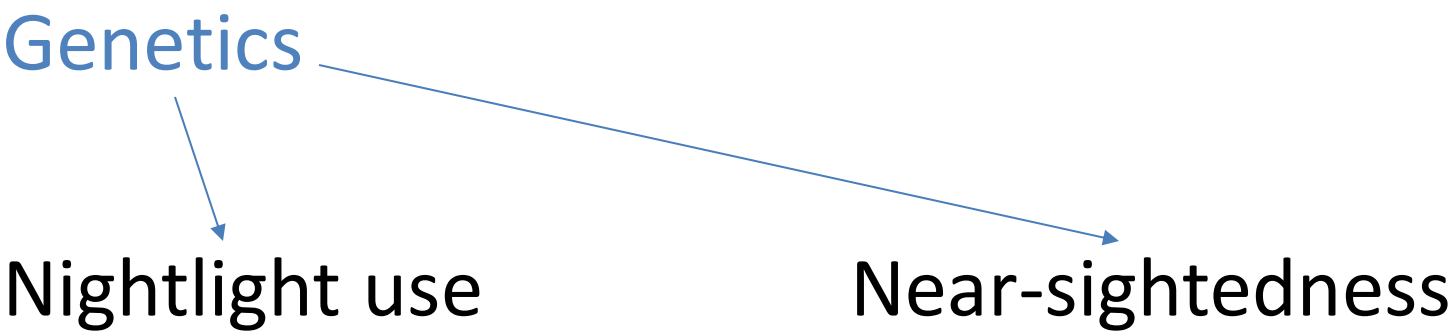
With a p-value of about 0, and assuming our samples are representative, we conclude that the proportion of near-sightedness increased when nightlights were used with infants.

*Would you say nightlight use with infants **caused** the increase in near-sightedness?*

No: children of parents who choose to use nightlights might be more genetically inclined to near-sightedness.

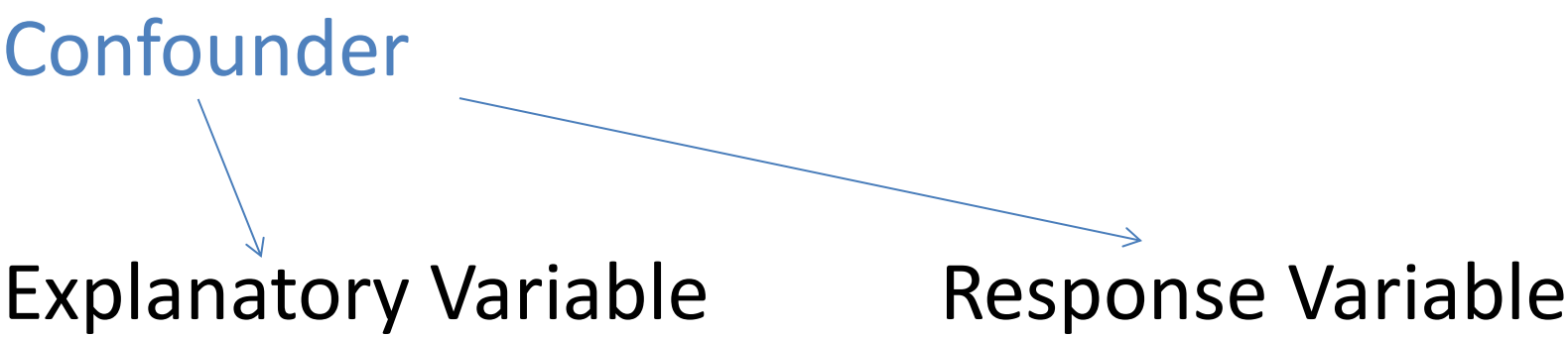
Confounding Variable

Example:



Confounding Variable

A third variable which influences both the explanatory and response variable, making it appear as though the explanatory variable influenced the response variable



Warning!

If we obtain random samples from our population, **and** get a statistically significant result (i.e. small p-value) we can **only** conclude our populations are different.

We have no information as to the *cause* of the difference due to the possibility of *confounding variables.*

Suggest a potential **confounding** variable for each study

- Does the size of the crowd (sold out or not) at a basketball game influence whether the home team wins or not?

RV = home team won (yes or no)

EV = size of crowd (sold out or not)

CV = ?

- Does taking a fish oil supplement reduce blood pressure?

RV = low blood pressure (yes or no)

EV = take fish oil (yes or no)

CV = ?

Suggest a potential **confounding** variable for each study

- Does the size of the crowd (sold out or not) at a basketball game influence whether the home team wins or not?

CV = fame of visiting team

EV = size of crowd (sold out or not)

RV = home team won (yes or no)

- Does taking a fish oil supplement reduce blood pressure?

CV = social economic status **or** general health

EV = take fish oil (yes or no)

RV = low blood pressure (yes or no)

Types of Studies

Observational Study: examine response variable in "naturally occurring" groups

Experimental Study: researcher assigns group membership, then examine response variable

Can an **experimental study** be carried out to answer these questions?

Or is the only choice an **observational study**?

- Does the size of the crowd (sold out or not) at a basketball game influence whether the home team wins or not?

RV = home team won (yes or no)

EV = size of crowd (sold out or not)

- Does taking a fish oil supplement reduce blood pressure?

RV = low blood pressure (yes or no)

EV = take fish oil (yes or no)

When can we identify a cause and effect relationship?

Sampling Method is	**Group Assignment is**	
	Random	Not random
Random		
Not random		

When can we identify a cause and effect relationship?