

Review: 3-3, 3-10

- Power
- Multi-test Correction: Bonferroni and Benjamini-Hochberg
- The Permutation Test
- “Tails”
- Regression goal and terminology
- Simple Linear Regression (Residual Sum of Squares)
- Multiple Linear Regression
- P-values for linear regression coefficients
- Logistic Regression: reweighted least squares

A lot can be answered with multiple linear regression

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im} + \epsilon_i$$

Mediation

Path Analyses (a type of “structured equation modeling”)

How much does **M** mediate the effect of **X** on **Y**?

$$Y = \beta_0 + c'X + bM + \epsilon_{c'b}$$

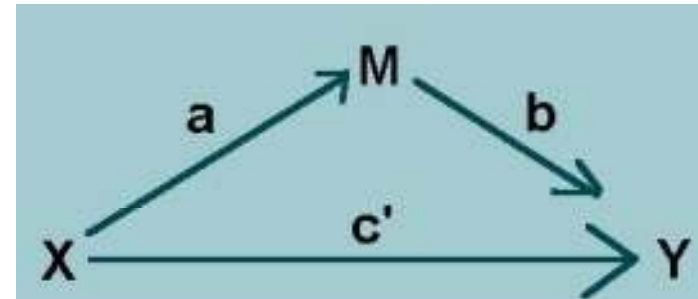
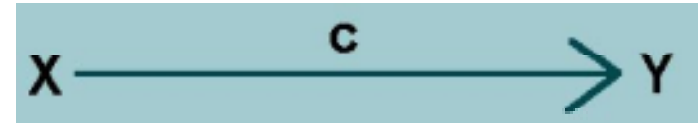
$$Y_i = \beta_0 + \text{beta}_1 X_{i1} + \text{beta}_2 X_{i2} + \dots + \text{beta}_m X_{im} + \epsilon_i$$

Mediation

Path Analyses (a type of “structured equation modeling”)

How much does **M** mediate the effect of **X** on **Y**?

$$Y = \beta_{0c} + cX + \epsilon_c$$
$$X = \beta_{0a} + aM + \epsilon_a$$
$$Y = \beta_{0c'} + c'X + bM + \epsilon_{c'b}$$



(Kenney, 2015)

<http://davidakenny.net/cm/mediate.htm>

Mediation

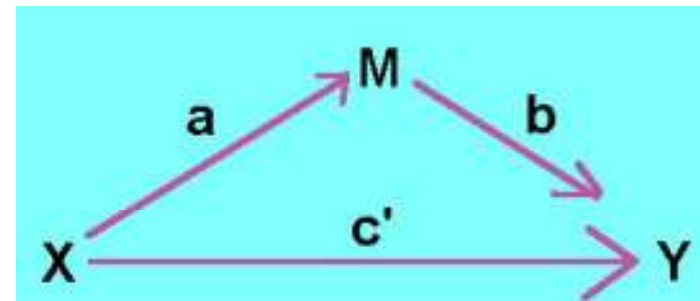
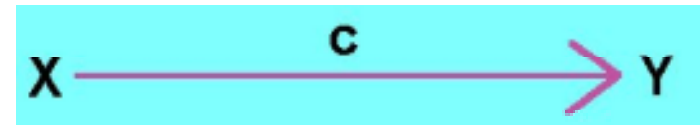
Path Analyses (a type of “structured equation modeling”)

How much does **M** mediate the effect of **X** on **Y**?

$$Y = \beta_{0c} + cX + \epsilon_c$$
$$X = \beta_{0a} + aM + \epsilon_a$$
$$Y = \beta_{0c'} + c'X + bM + \epsilon_{c'b}$$

Effect size: often reported as $c - c'$.

Used for *basic* causal inference.

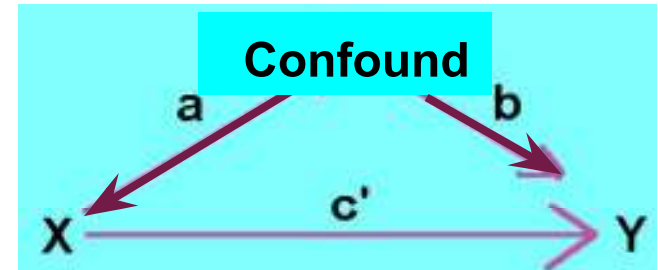


(Kenney, 2015)

<http://davidakenny.net/cm/mediate.htm>

Hierarchical Linear Models (HLM)

Problem: Sometimes variables at one unit of analysis are *confounded* by a variable at another level.



Hierarchical Linear Models (HLM)

Problem: Sometimes variables at one unit of analysis are *confounded* by a variable at another level.

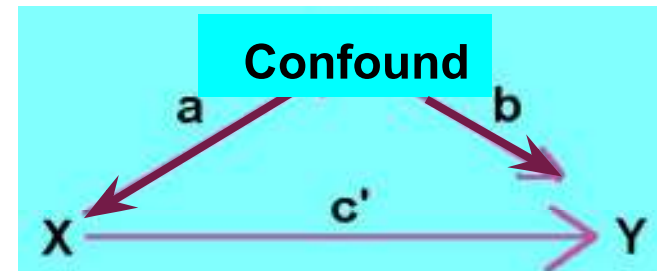
Examples:

Pot heads are more likely to say “hella”
but really californians are more like to say “hella” and be potheads.

X = use of “hella”

Y = pot-head or not

Confound = from california?



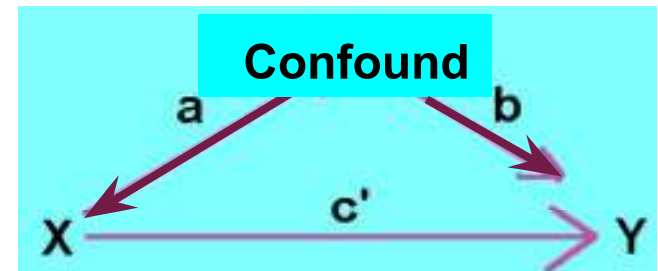
Hierarchical Linear Models (HLM)

Problem: Sometimes variables at one unit of analysis are *confounded* by a variable at another level.

Examples:

Pot heads are more likely to say “hella”
but really californians are more like to say “hella” and be potheads.

Females are more likely to post pictures of food
but really both food posts and females are more common on Pinterest.



Hierarchical Linear Models (HLM)

Problem: Sometimes variables at one unit of analysis are *confounded* by a variable at another level.

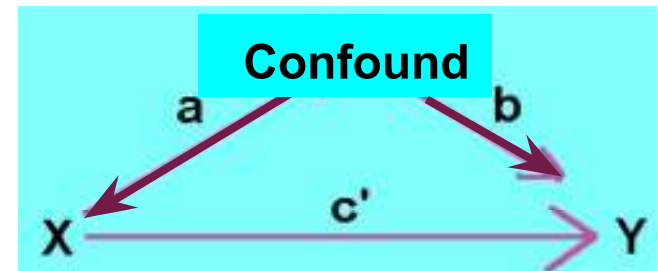
Examples:

Pot heads are more likely to say “hella”
but really californians are more like to say “hella” and be potheads.

Females are more likely to post pictures of food
but really both food posts and females are more common on Pinterest.

Solution: include aggregate confounding variable as a covariate in multiple linear regression. (also useful for prediction)

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im} + \epsilon_i$$



Hierarchical Linear Models (HLM)

Problem: Sometimes variables at one unit of analysis are *confounded* by a variable at another level.

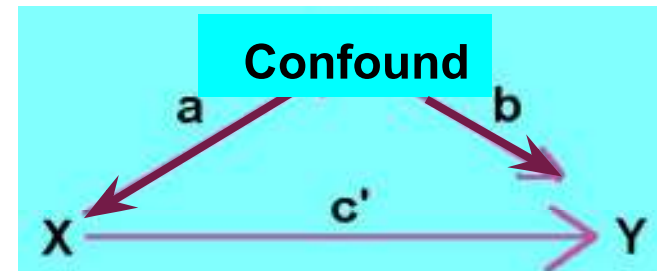
Examples:

Pot heads are more likely to say “hella”
but really californians are more like to say “hella” and be potheads.

Females are more likely to post pictures of food
but really both food posts and females are more common on Pinterest.

Solution: include aggregate confounding variable as a covariate in multiple linear regression. (also useful for prediction)

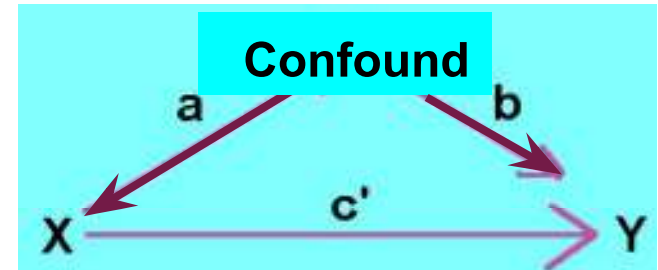
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_m X_{im} + \epsilon_i$$



Hierarchical Linear Models (HLM)

Problem: Sometimes variables at one unit of analysis are *confounded* by a variable at another level.

$$Y = \beta_0 + \beta_1 X_1 + \beta_A A + \epsilon$$



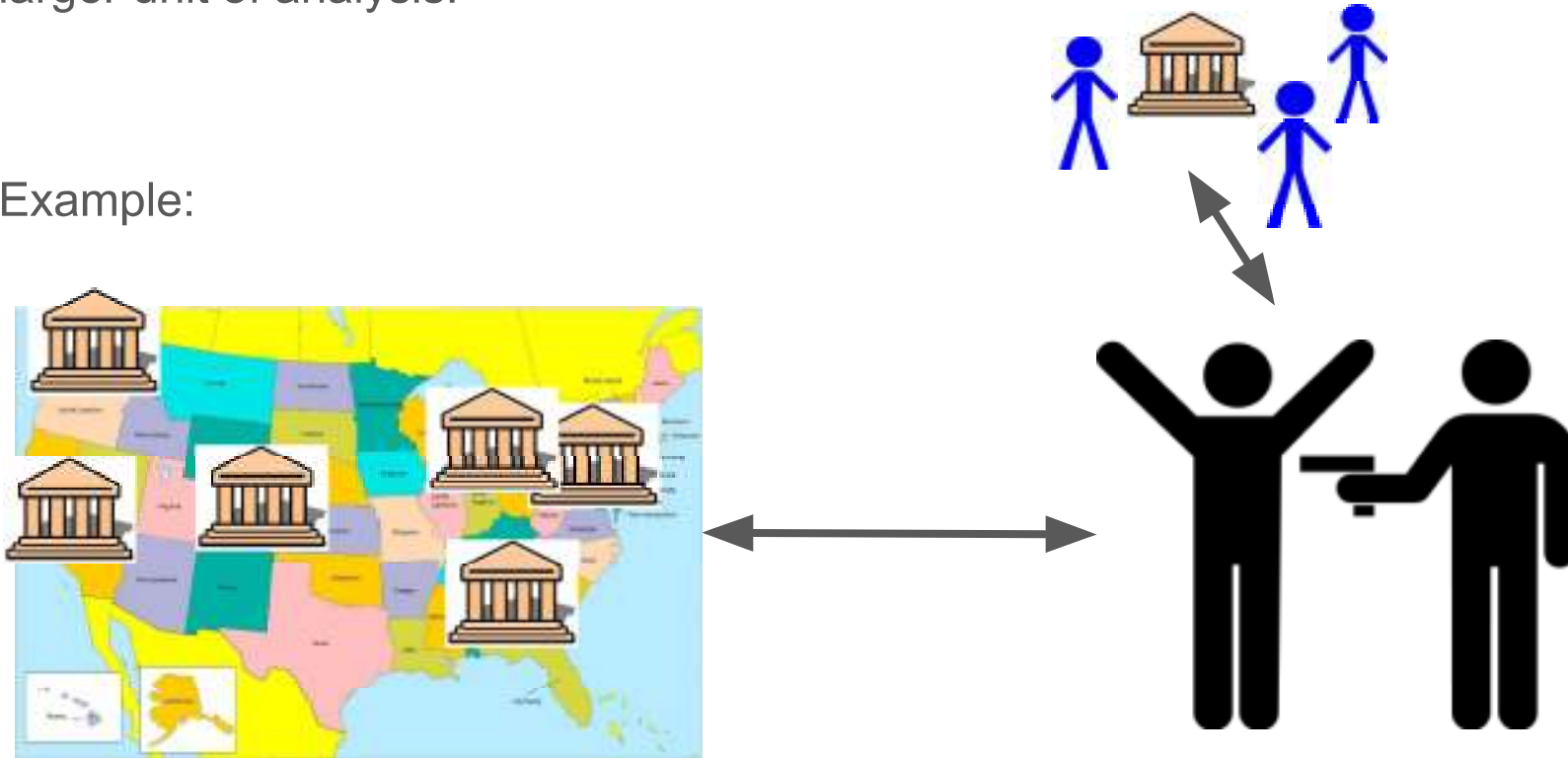
A : aggregate indicator variable (is in region or not? Pinterest usage).

Solution: include aggregate confounding variable as a covariate in multiple linear regression. (also useful for prediction)

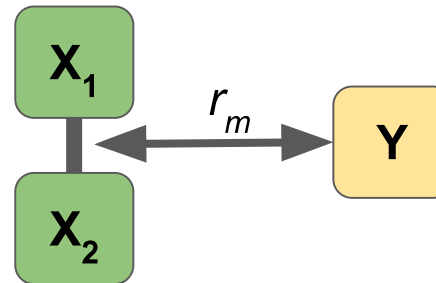
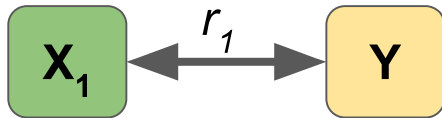
Ecological Fallacy

The assumption that an effect at one unit of analysis will hold for a smaller or larger unit of analysis.

Example:



Moderation (interaction)



When $r_1 \neq r_m$, X_2 moderates the relationship between X_1 and Y .

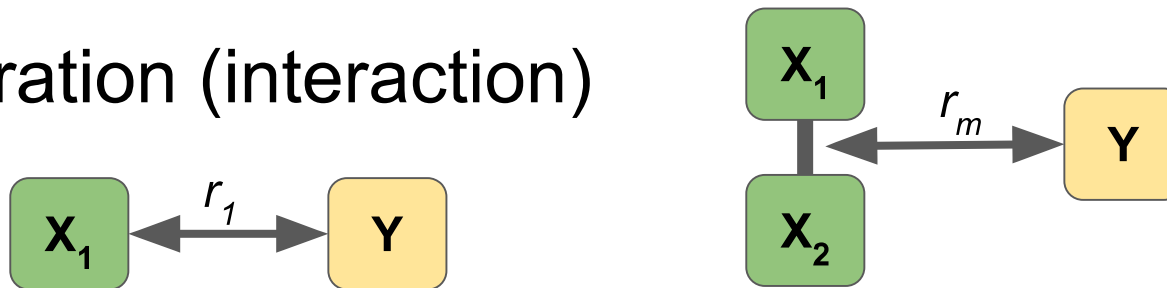
Examples:

Y : Attend church? X_1 : Agreeableness, X_2 : From US?

Movie Reviews:

Y : Rated Depressing, X_1 : "death" in review, X_2 : Silly Horror Movie?

Moderation (interaction)



When $r_1 \neq r_m$, X_2 moderates the relationship between X_1 and Y .

More precisely moderation analyses fit the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_M X_1 X_2 + \beta_2 X_2 + \epsilon$$

$X_1 X_2$: The interaction term.

β_M can then be tested for significance using the same t-test we use for any individual coefficient in multiple linear regression