announcements

- welcome back!
- Problem Set 7 is due on Wednesday
- there will be drill this week
  - it will cover analysis of continuous-by-group interactions and basic mediation analysis

1

general exam feedback

- be thorough in interpreting interactions

> **The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant**
>
> Andrew Gelman & Hal Stern

- be informative in interpreting statistical significance: don't simply say that one effect is significant and the other isn't significant ... note the difference between significant and not significant, whether, etc.

	trash	control
positive	5.5	6.4
negative	5.6	5.4

- power analysis is hard; please check the answer key re: 3g & ask questions if you have them

2

announcements

- Problem Set 7 is due on Wednesday
- there will be drill this week
  - it will cover analysis of continuous-by-group interactions and basic mediation analysis

3

ANCOVA: last words

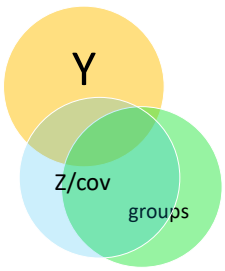
March 25, 2024

4

first: problems with overlap between covariate & groups

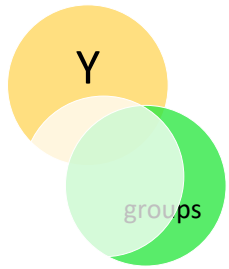
5

conceptually



6

conceptually



7

what is the problem?

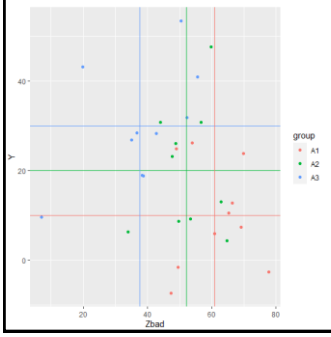
- with correlated predictors (i.e., tolerance < 1), giving credit for overlapping variance explained is complicated
- depends on *causal priority*; which predictor influences the outcome first
- ANCOVA, when done sequentially, assumes the covariate influences the outcome before the grouping variable does
- if this is incorrect, interpreting group differences controlling for a covariate is fraught w/difficulty
- as Cohen & Cohen put it (with my slight edits for provinciality), *the difference in mean height between the Himalayan and Ozark mountains, adjusting for differences in atmospheric pressure, is about zero*

8

sensitive content: CSA

9

problematic data w/actual *M*s



these groups differ with respect to *Y*

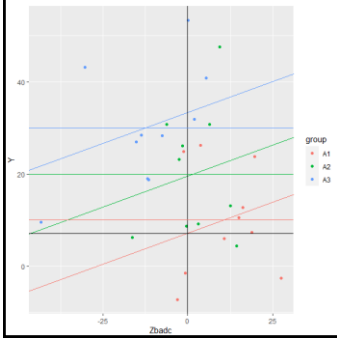
$M_{A1} = 10$
$M_{A2} = 20$
$M_{A3} = 30$

these groups also differ with respect to Zbad

$M_{A1} = 60.9$
$M_{A2} = 52.2$
$M_{A3} = 37.7$

10

problematic data w/actual *M*s



adjusted means

$M_{A1} = 7.1$
$M_{A2} = 19.5$
$M_{A3} = 33.4$

11

special design issue: pretest-posttest

- imagine we're interested in comparing the effectiveness of two methods of teaching reading
- at the beginning of a school year, we give students a standardized test; call this variable Z
- students are randomly assigned to learn to read by one of the two methods; call this variable X
- at the end of the year, the students take the same standardized test; call this variable Y
- how should we analyze this?

12

we have options: first, change scores ($Y - Z$)

- the model for this would be

$$Y_i - Z_i = \beta_0 + \beta_1 X_1 + \varepsilon_i$$

- rearranging this by moving *Z* to the right side

$$Y_i = \beta_0 + \beta_1 X_1 + Z_i + \varepsilon_i$$

- this implies that the slope of *Z* is 1; it's not an estimated parameter

$$Y_i = \beta_0 + \beta_1 X_1 + 1Z_i + \varepsilon_i$$

13

we have options: second, an ANCOVA

- the model for this would be

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 Z_i + \varepsilon_i$$

- because we've added a parameter (β_2) instead of setting it equal to 1, this will give us a better fit
- if you have change scores, do an ANCOVA w/pretest scores as a covariate
- the main exception is if the β_2 estimate is ≈ 1, then the 1 *df* cost to estimate it might not be worth it
- if you work in an area where change scores are commonly used, read around to see how others handle them

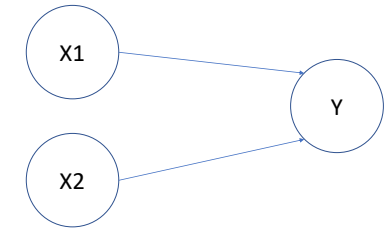
14

other third-variable patterns (covariates, confounds, mediators, etc.)

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3689437

15

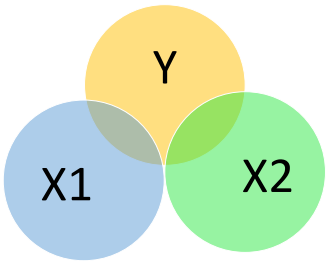
covariates, confounds, mediators



https://journals.sagepub.com/doi/full/10.1177/2515245917745629

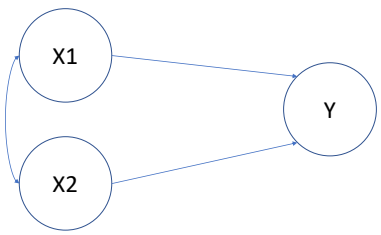
16

X2 is a “covariate”



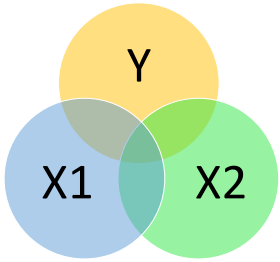
17

covariates, confounds, mediators



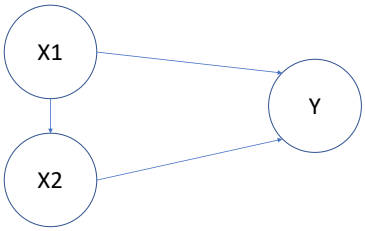
18

X2 is a (partial) confound



19

covariates, confounds, mediators

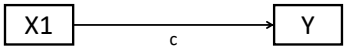


20

mediation

(Baron & Kenny, 1986; Hayes, 2009; Rucker et al., 2011)

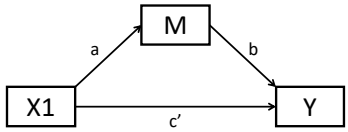
- if one variable influences another through an intervening variable, the intervening variable is typically called a *mediator*



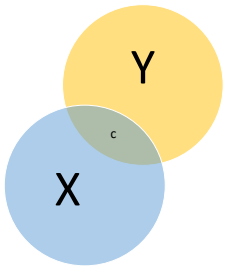
21

mediation

- if one variable influences another through an intervening variable, the intervening variable is typically called a *mediator*

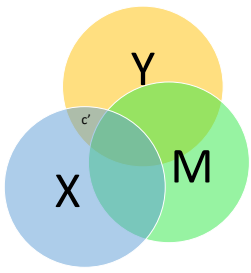


22



23

M is a mediator



24

total = direct + indirect

- total effect of X on Y = *c*
- direct effect of X on Y = *c'*
- indirect effect of X on Y via M = *ab*
- *c* = *c'* + *ab*
- *ab* = *c* – *c'* (the indirect effect = total – direct)

25

Baron & Kenny's causal steps approach

1) regress Y on X: *c* must be significant
2) regress M on X: *a* must be significant
3) regress Y on X & M: *b* must be significant
   - if *c* > *c'* and *c'* is significant → "partial mediation"
   - if *c* > *c'* and *c'* is NS → "full mediation"

But this approach has low power!

26

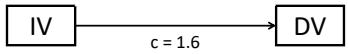
an example

- adolescents diagnosed with bipolar disorder are randomly assigned to a treatment group (a family counseling intervention + the usual pharmaceutical regimen) or a control group (only the pharmaceutical)
- the outcome is a measure of symptoms taken at 8 weeks after treatment begins
- we suspect that the counseling will be effective by reducing criticism; this is measured at 7 weeks

27

model 1: symptoms ~ treatment

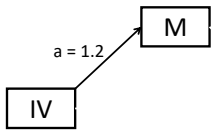
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.1000	0.2739	18.623	3.29e-13
X	1.6000	0.5477	2.921	0.00912



28

model 2: criticism ~ treatment

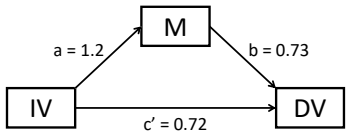
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1000	0.1871	16.570	2.41e-12
X	1.2000	0.3742	3.207	0.00489



29

model 3: symptoms ~ tx + crit

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.8365	0.9847	2.881	0.0104
X	0.7238	0.6123	1.182	0.2535
C	0.7302	0.3077	2.373	0.0297



a*b = 1.2*0.73 = 0.876 = c – c' = 1.6 – 0.724 = 0.876 ✔

30

enough

a little more about mediation on Wednesday

31