more about ANCOVA

March 13, 2024

1

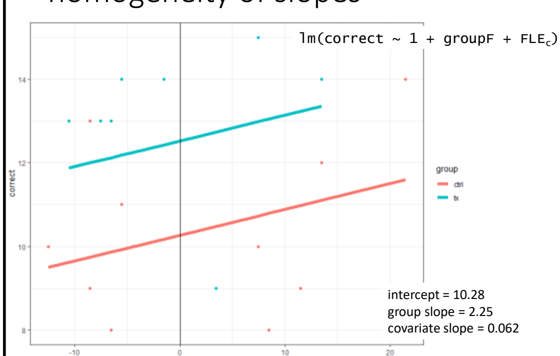
ANCOVA comes w/baggage

- the typical ANCOVA model does not include a group-by-covariate interaction
- that is, it assumes the covariate-outcome relationship (i.e., slope) is the same for every group

- the typical ANCOVA is easier to interpret if the covariate and groups are independent
- interpretation is more difficult if not

2

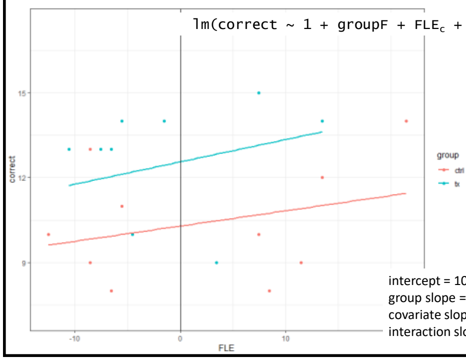
homogeneity of slopes



3

allowing slope to vary

`lm(correct ~ 1 + groupF + FLE_c + group:FLE_c)`



intercept = 10.29
group slope = 2.27
covariate slope = 0.054
interaction slope = 0.025

4

what about the independence of covariate and groups?

- this is all but guaranteed with random assignment
- how to test? model the *covariate as the outcome* and group(s) as the predictor
- you want this to be non-significant!

5

what if the covariate and groups are related?

- what if you don't have random assignment (or if you have bad luck)?
- interpretation is complicated!
- please read Miller & Chapman (2001); it's so good

6

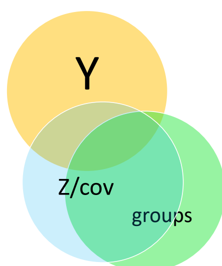
group (A) – covariate (Z) related: complications

an example

- Y = yield in corn plants
- A = two varieties of corn (blue vs white)
- result: white > blue
- but white is taller than blue (Z = height)
- regress Y on Z, then include A, no effect of A
- what should we conclude?

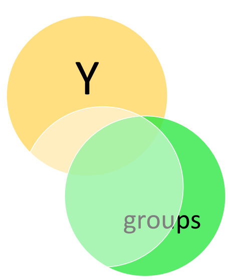
7

conceptually



8

conceptually



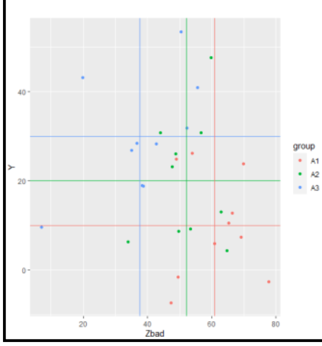
9

what is the problem?

- with correlated predictors (i.e., tolerance < 1), giving credit for overlapping variance explained is complicated
- it depends on causal priority; which predictor influences the outcome first
- recall that the ANCOVA can be done as a sequential analysis
- but this assumes that the covariate influences the outcome before the grouping variable does
- if this assumption is incorrect, interpreting group differences after controlling for the covariate is fraught with difficulty

10

problematic data w/actual *M*s



these groups differ with respect to *Y*

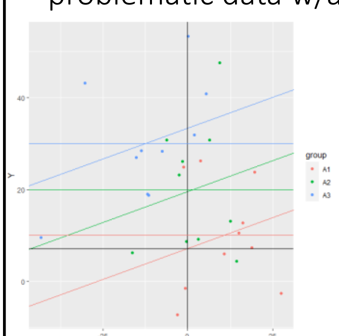
M_{A1} = 10
M_{A2} = 20
M_{A3} = 30

these groups also differ with respect to Zbad

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

11

problematic data w/actual *M*s



adjusted means

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

12

if the covariate interacts w/the predictor of interest

- try to understand why!
- in the words of one textbook: “One would simply live with the more complex model and interpret the resulting significant covariate × condition interactions.” 😂

13

sometimes the cov × group interaction is of interest

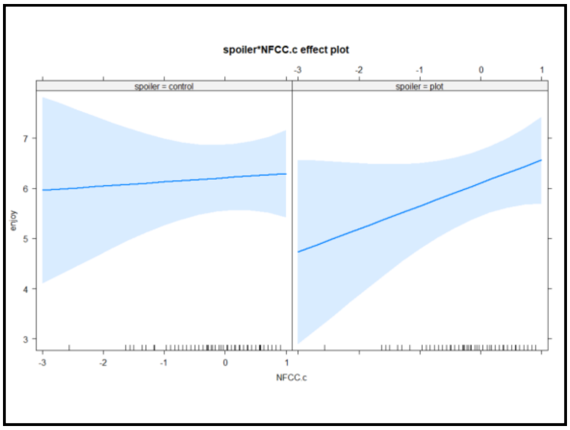
- that is, the goal of using the covariate is not to gain power for comparing groups
- instead, one might expect the covariate to interact with a grouping variable (whether in an experiment or in a quasi-experiment)
- in this case, the interaction is of theoretical interest and should be modeled to estimate the parameter(s) of interest

14

an example (from my research; sorry)

- we hypothesized that spoilers given before reading a short story would generally reduce enjoyment
- but we thought that the effect of spoilers would be different among those low vs high in “need for closure”
- those low in need for closure might be bothered by the spoilers
- those high in need for closure might appreciate the spoilers
- this is a grouping (spoiler) × covariate interaction

15



16

special design issue: pretest-postest

- 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?

17

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

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

18

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

19