# a little more multilevel modeling

April 22, 2024

1

## two classrooms data

- one predictor (X, a pretest), mean-centered
- one outcome (Y, a standardized test)
- one class scores 10 points higher than average than the other
- the X-Y relationship in the two classes is the same

2

## Option 3

- Modeling not only the effect of the pretest at the subject level
- Also modeling the differences in classrooms

$$\hat{Y}_i = \hat{\beta}_{0,j} + \beta_1 X_{pretest,i} + e_i$$
$$\hat{\beta}_{0,j} = \gamma_0 + u_j$$

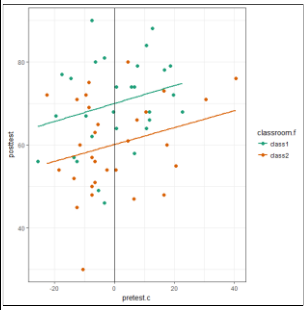
- This is called a random-intercept model, and can be presented as one equation

$$\hat{Y}_i = [\gamma_0 + \beta_1 X_{pretest,i}] + [u_j + e_i]$$

3

## exploring the twoclassrooms data

# Section 1




- things to notice:
  - classes have similar slopes
  - class1 has an intercept of 70
  - class2 has an intercept of 60
- think of intercepts as DV means (when predictors are centered)

4

## separate regressions

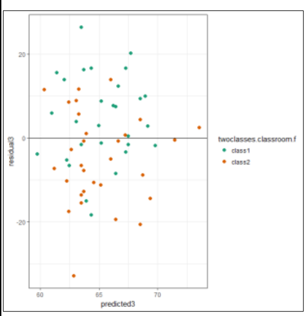
# Section 2

- class 1 regression equation (rounded)
  $\hat{Y} = 70 + 0.2 * pretest$
- class 2 regression equation (rounded)
  $\hat{Y} = 60 + 0.2 * pretest$

5

## analysis option: collapsing across classrooms

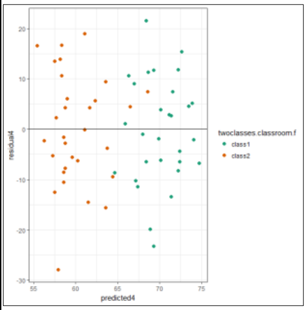
# Section 3




- $R^2$ = .055; $s_{resid}$ = 11.9
  $\hat{Y} = 65 + 0.2 * pretest$
- intercept between 60 & 70
- pro: $n$ = 60 better than $n$ = 30
- con: independence assumption is in bad shape
  - residuals are clustered (class1 mostly > 0, class2 mostly < 0)
  - residuals $M_{class1}$ = 4.9
  - residuals $M_{class2}$ = -4.9

6

## analysis option: modeling classrooms as a predictor
`# Section 4`




- $R^2$ = .22; $s_{resid}$ = 10.9

$\hat{Y} = 70 + 0.2pre + -10class$

- intercept represents class1
- class slope represents class2 difference from class1
- pro: independence of residuals is in good shape
  - residuals not clustered
  - residuals $M_{class1}$ = 0
  - residuals $M_{class2}$ = 0
- con: scaling up (e.g., to 30 classes) costs parameters (which we probably don't care about)

7

## How to run a multilevel model using `lmer()` in the lme4 package

```
lmer(data = twoclasses,
   posttest ~ (1 | classroom.d) + pretest.c )
```

8

## Parsing part of the `summary()` of an `lmer()` call

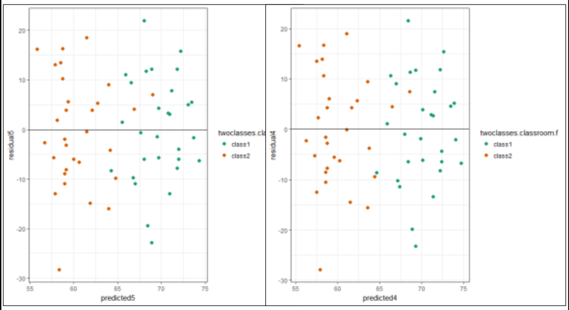
```
Random effects:
 Groups      Name        Variance Std.Dev.
 classroom.d (Intercept)  44.72    6.688
 Residual                118.52   10.887
Number of obs: 60, groups:  classroom.d, 2

Fixed effects:
            Estimate Std. Error t value
(Intercept) 65.0333     4.9333  13.182
pretest.c    0.2089     0.1036   2.017
```

9

## Analysis option: multilevel modeling

*# Section 5*



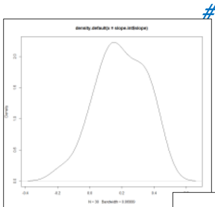

10

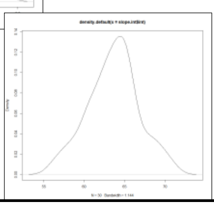
## scaling up to 30 classes

11

## 30 (!) separate regressions

*# Section 6*





- 30 regression equations
  - class1: $\hat{Y} = 66.0 + (0.40)\mathit{pretest}$
  - class2: $\hat{Y} = 70.4 + (-0.06)\mathit{pretest}$
  - class3: $\hat{Y} = 60.4 + (0.25)\mathit{pretest}$
  - …
- all slopes: *M* = 0.19, *SD* = 0.15
- all intercepts: *M* = 63.6, *SD* = 3.17

12

## Analysis option: modeling classrooms as a predictor

`#  Section 7`

$$\hat{Y} = 65.9 + 0.2pre + (-2.4)class10 + (2.4)\text{class}11 + \cdots$$

- intercept represents class1
- each class now has a slope; there are 29 slopes
- we can add each slope to the intercept to find one for each class
- do we care about each intercept?
- if not, why not just model how much variability there is in them?

13

## Analysis option: MLM

`#  Section 8`

```
Random effects:
 Groups   Name        Variance Std.Dev.
 class.id (Intercept)   4.517   2.125
 Residual             148.925  12.203
Number of obs: 822, groups:  class.id,
30

Fixed effects:
            Estimate Std. Error t value
(Intercept) 63.58769    0.57648  110.30
pretest.c    0.20042    0.03509    5.71
```

$$\hat{Y} = 63.6 + 0.2pre$$

- intercept represents everyone's mean
- variability in intercepts
  var = 4.52 (*SD* = 2.15)
- one slope for everyone (we're assuming it's fixed for everyone, so we get no estimate of variability)

14

## the model we're estimating

$$Y_{ij} = [\gamma_{00} + \beta_1 X_{ij}] + [u_{0j} + e_{ij}]$$

<u>Translated</u>

($Y$) each student's posttest score (is a function of) =
($\gamma$) the grand mean of all scores +
($X$) the student's pretest score (times slope) +
($u$) the student's classroom's residual +
($e$) the student's residual

15

## lmer output (for 30 classes)

```
Linear mixed model fit by REML ['lmerMod']
Formula: posttest ~ pretest.c + (1 | class.id)
   Data: data

REML criterion at convergence: 6465.8

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.84671 -0.74761  0.01867  0.63232  3.09975

Random effects:
 Groups   Name        Variance Std.Dev.
 class.id (Intercept)   4.517   2.125
 Residual             148.925  12.203
Number of obs: 822, groups:  class.id, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept) 63.58769    0.57648  110.30
pretest.c    0.20042    0.03509    5.71
```

16