Multilevel models, part 2

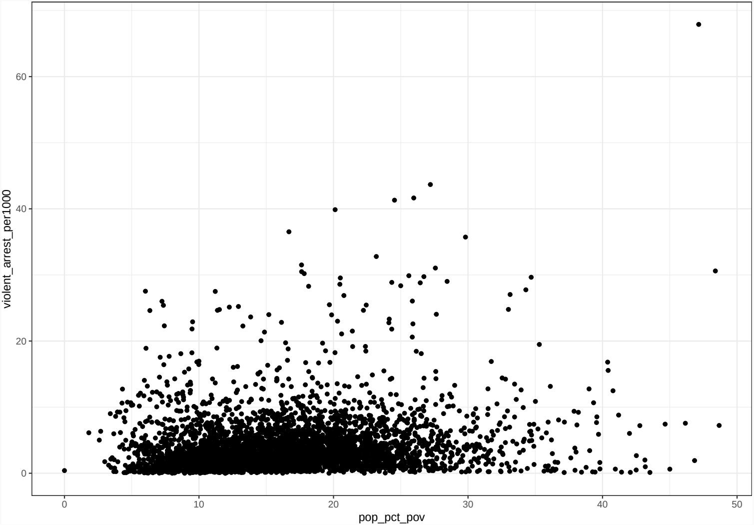
Frank Edwards

Arrests and poverty

Data for today (available at `https://github.com/f-edwards/intermediate_stats/tree/master/data`)

```
dat<-read_csv("./data/violent_arrest_data.csv")
```

The Distribution of poverty and crime across counties



A simple model

This model, with no included parameters for group differences, is called a *complete pooling* model. Data from all groups is used to estimate a single intercept and single slope.

Where violent crime rates are y, poverty rates are x, counties are c, states are s

$$y_c = \beta_0 + \beta_1 x_c + \varepsilon_c$$

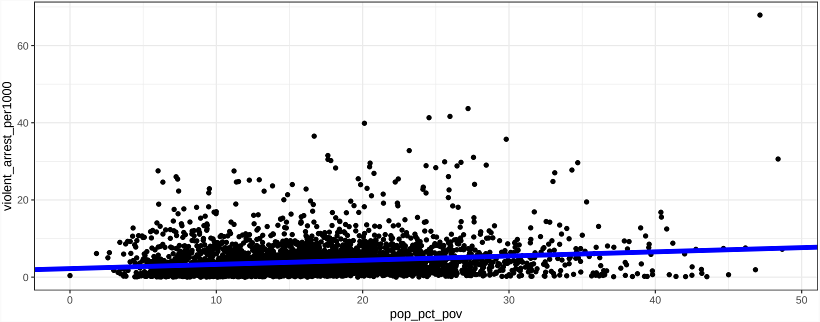
$$\varepsilon \sim N(0, \sigma^2)$$

Estimating the complete-pooling model

```r
m_complete_pooling <- lm(violent_arrest_per1000 ~ pop_pct_pov,
                         data = dat)
broom::tidy(m_complete_pooling)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)    2.21    0.171        12.9 1.74e-37
## 2 pop_pct_pov    0.109   0.00963      11.3 4.28e-29
```

Visualizing the model



Let's estimate a model where NO information across states is included in our estimation of intercepts.

- Every state gets their own intercept (computed as the exact within-state violent crime rate mean).

Estimating the state fixed effects model

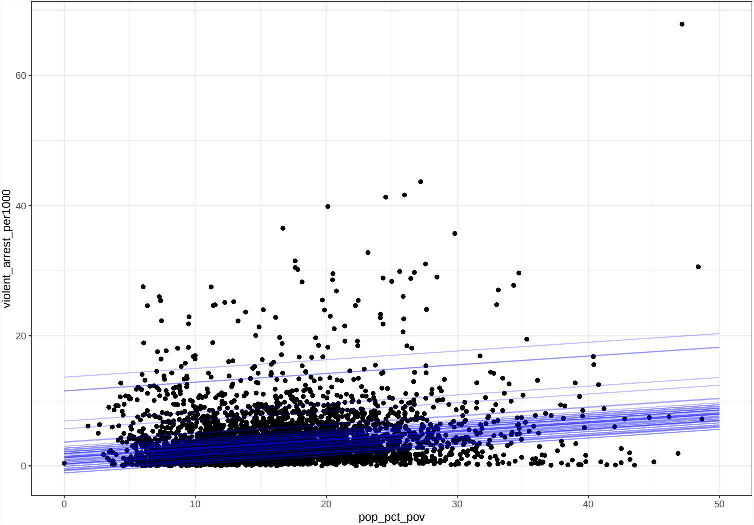
```
m_state_fe<-lm(violent_arrest_per1000 ~
                 pop_pct_pov +
                 factor(stusps),
               data = dat)
broom::tidy(m_state_fe)
```

```
## # A tibble: 52 x 5
##    term              estimate std.error statistic  p.value
##    <chr>                <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)         11.6      0.808     14.3   2.78e-45
##  2 pop_pct_pov          0.134    0.0103    12.9   1.24e-37
##  3 factor(stusps)AL    -9.07     0.909     -9.98  3.35e-23
##  4 factor(stusps)AR    -9.62     0.903    -10.7   3.58e-26
##  5 factor(stusps)AZ   -11.9      1.11     -10.8   7.64e-27
##  6 factor(stusps)CA   -10.2      0.888    -11.4   7.76e-30
##  7 factor(stusps)CO   -10.5      0.898    -11.7   3.56e-31
##  8 factor(stusps)CT   -11.2      1.55      -7.22  6.30e-13
##  9 factor(stusps)DC     2.09     2.78       0.751 4.53e- 1
## 10 factor(stusps)DE   -12.3      2.32      -5.29  1.30e- 7
## # i 42 more rows
```

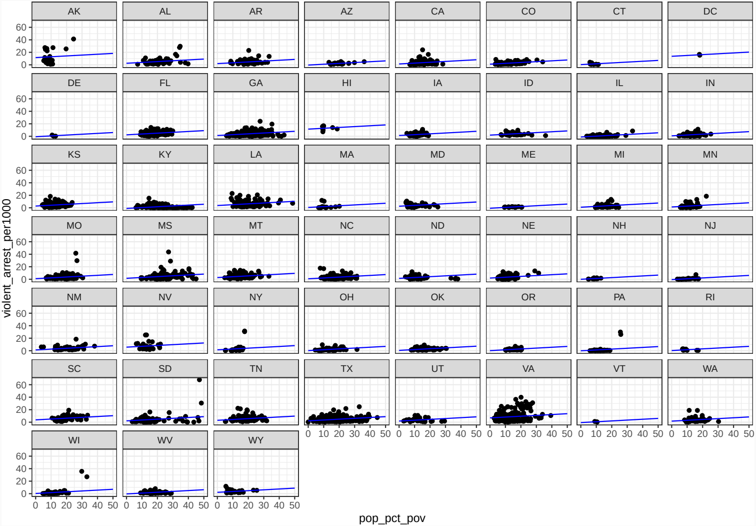
What an intercept means in practice

Intercepts are computed as the average violent arrest rate in each state, when poverty is fixed at zero - State intercepts adjust for differences in average violent arrests rates at the county-level *across states*

Visualizing the fixed effects (state intercepts) model



Another visualization



Do we think that the relationship between poverty and arrest rates is identical across states?

- Intercepts capture *average* arrest levels within the state. They adjust for variation in arrest rates for both high and low poverty counties
- Slopes capture *differences* in arrest rates between high and low poverty counties both within and across states.
- What processes could account for variation in *intercepts* and/or *slopes* in our study?

Varying slopes

Returning to the fixed effects (intercepts) model

Where y is a violent arrest rate, x is a poverty rate, c indicates a county, and s indicates a state

$$y_c = \beta_{0s} + \beta_1 x_c + \varepsilon_c$$

What does this model do?

Modifying the model to allow for varying slopes

Where y is a violent arrest rate, x is a poverty rate, c indicates a county, and s indicates a state

Fixed effects intercepts

$$y_c = \beta_{0s} + \beta_1 x_c + \varepsilon_c$$

Fixed effects intercepts AND slopes

$$y_c = \beta_{0s} + \beta_{1s} x_c + \varepsilon_c$$

Breaking down how the fixed effects interaction model works

We are now effectively adding an *interaction* for the poverty variable and the state variable.

We can rewrite the regression model like this, where β_s is the adjustment we add to each state's slope from the reference category (β_2).

$$y_c = (\beta_0 + \beta_{1s}) + (\beta_2 + \beta_{3s})x_c + \varepsilon_c$$

Now our relationship between poverty (x) and arrest rates y depends both on poverty levels AND on the state a county is in. That's an interaction.

Breaking down how the fixed effects interaction model works

We are now effectively adding an *interaction* for the poverty variable and the state variable.

We can rewrite the regression model like this, where β_s is the adjustment we add to each state's slope from the reference category (β_2).

$$y_c = (\beta_0 + \beta_{1s}) + (\beta_2 + \beta_{3s})x_c + \varepsilon_c$$

Now our relationship between poverty (x) and arrest rates y depends both on poverty levels AND on the state a county is in. That's an interaction.

We've redefined our slope parameter to $\beta_2 + \beta_{3s}$

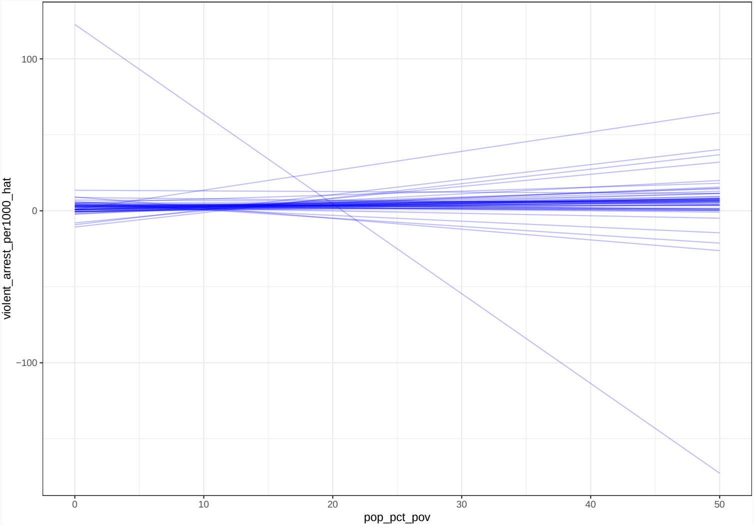
Estimating the fixed effects interaction model in R

```
m_state_fe_int<-lm(violent_arrest_per1000 ~
                 pop_pct_pov *
                 factor(stusps),
               data = dat)
broom::tidy(m_state_fe_int)
```

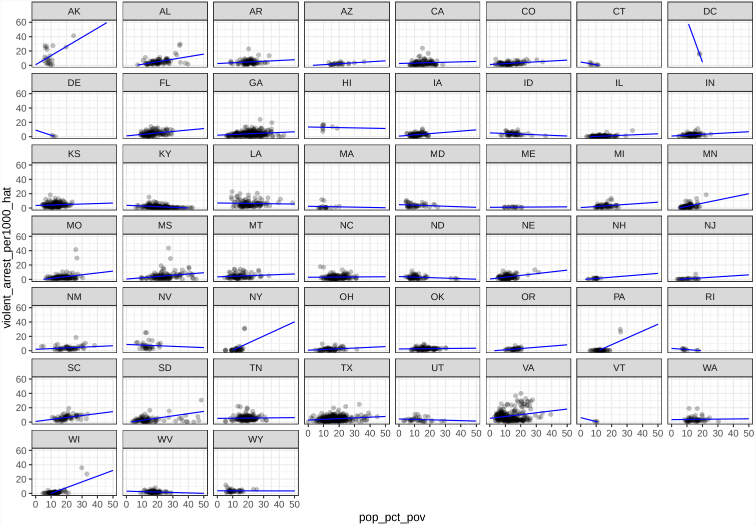
```
## # A tibble: 102 x 5
##    term              estimate std.error statistic  p.value
##    <chr>                <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)          0.839     1.84      0.457 6.48e- 1
##  2 pop_pct_pov          1.28      0.177     7.20  7.20e-13
##  3 factor(stusps)AL    -3.06      2.32     -1.32  1.87e- 1
##  4 factor(stusps)AR     1.66      2.53      0.657 5.11e- 1
##  5 factor(stusps)AZ    -1.22      3.46     -0.354 7.23e- 1
##  6 factor(stusps)CA     1.76      2.20      0.797 4.25e- 1
##  7 factor(stusps)CO     0.361     2.07      0.174 8.62e- 1
##  8 factor(stusps)CT     3.78      5.67      0.666 5.06e- 1
##  9 factor(stusps)DC   122.      407.        0.299 7.65e- 1
## 10 factor(stusps)DE     8.30     27.4       0.303 7.62e- 1
## # i 92 more rows
```

How many betas do we have (so many betas...)????

What the model is doing



What the model is doing



Three approaches to pooling data

- Under a complete-pooling model, we estimate a single intercept *and slope* for the full data based on the average violent crime rate *across all counties*
- Under a no-pooling model (the fixed effects interaction model), we estimate an intercept AND slope for every state based *only* on the data from that state
- Under a partial-pooling model, we estimate an intercept AND slope for every state based based on a model that *assumes that state intercepts AND slopes are generated from an underlying probability distribution*

A basic multilevel model (intercepts only)

Where violent crime rates y are a function of poverty rates x in state s, and county c

$$y_c = (\beta_0 + \alpha_s) + \beta_1 x_c + \varepsilon_c$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\alpha \sim N(0, \sigma_\alpha^2)$$

A A basic multilevel model (intercepts and slopes)

$$y_c = (\beta_0 + \alpha_s) + (\beta_1 + \delta_s)x_c + \varepsilon_c$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\alpha \sim N(0, \sigma^2_\alpha)$$

$$\delta \sim N(0, \sigma^2_\delta)$$

Estimating the random intercepts and slopes model

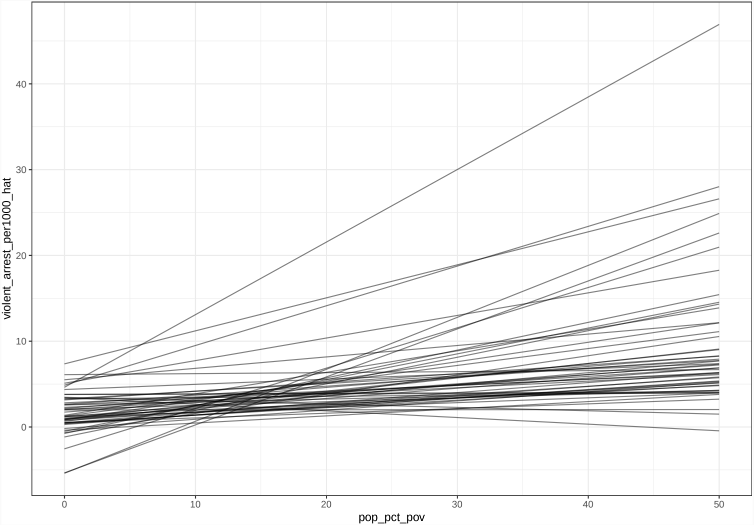
```r
library(lme4)
# RE specification gets complex
# google lmer cheat sheet for support
m_state_ml<-lmer(violent_arrest_per1000 ~
                   pop_pct_pov +
                   (1|stusps) + #intercepts for each state
                   (0 + pop_pct_pov|stusps), #slopes for each state, uncorrelated
                 data = dat)
```

Model output

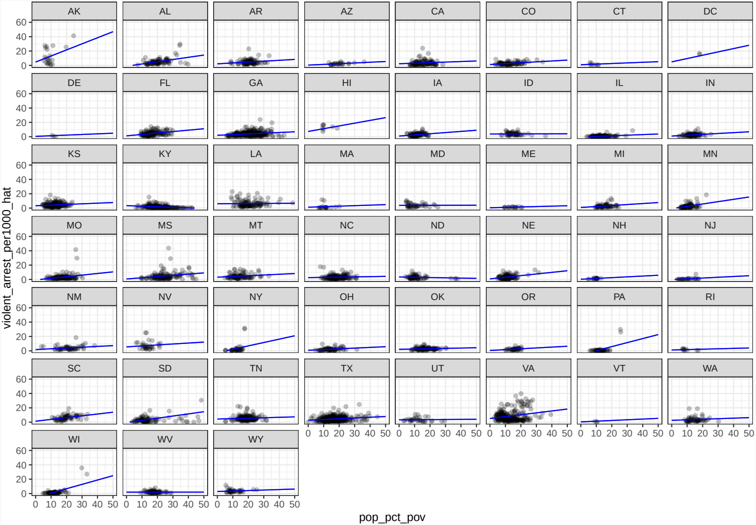
```r
summary(m_state_ml)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## violent_arrest_per1000 ~ pop_pct_pov + (1 | stusps) + (0 + pop_pct_pov |
##     stusps)
##    Data: dat
##
## REML criterion at convergence: 23566
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4228 -0.4845 -0.1539  0.2604 14.8020
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  stusps   (Intercept)  7.67470 2.7703
##  stusps.1 pop_pct_pov  0.03957 0.1989
##  Residual             13.41472 3.6626
## Number of obs: 4286, groups:  stusps, 51
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.62472    0.45475   3.573
## pop_pct_pov  0.16157    0.03261   4.954
##
## Correlation of Fixed Effects:
##             (Intr)
## pop_pct_pov -0.239
```

What the model does



What the model does

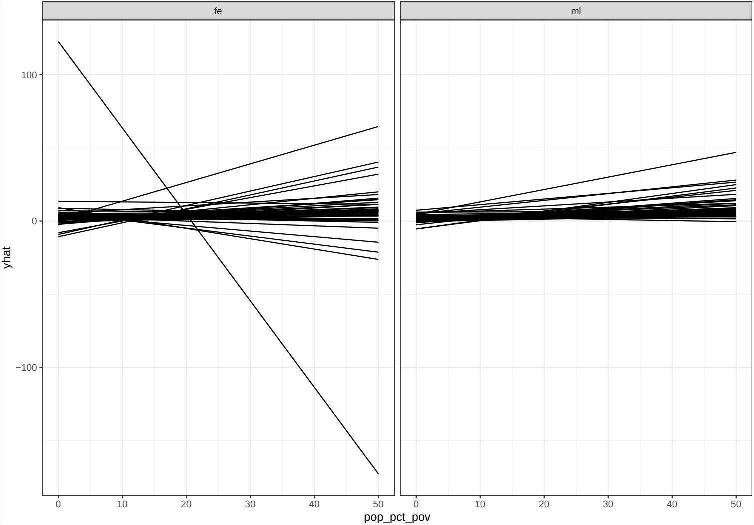


Let's compare multilevel and fixed effects model fits

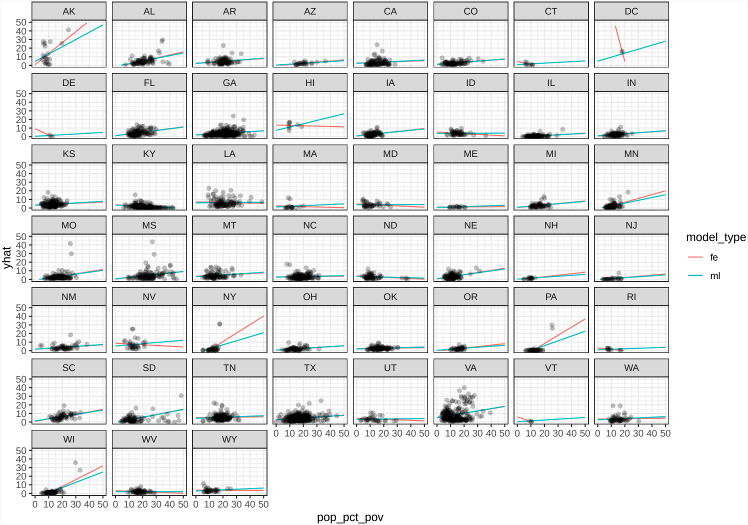
```r
fake_data<-expand_grid(stusps = unique(dat$stusps),
                       pop_pct_pov = 0:50)
# generate yhat for both fe and ml models
fake_data<-fake_data %>%
  mutate(fe = predict(m_state_fe_int, newdata = fake_data),
         ml = predict(m_state_ml, newdata = fake_data))
# format for plotting
fake_data<- fake_data %>%
  pivot_longer(cols = fe:ml,
               names_to = "model_type",
               values_to = "yhat")

head(fake_data)
```

```
## # A tibble: 6 x 4
##   stusps pop_pct_pov model_type   yhat
##   <chr>        <int> <chr>       <dbl>
## 1 AL               0 fe         -2.22
## 2 AL               0 ml         -1.16
## 3 AL               1 fe         -1.87
## 4 AL               1 ml         -0.853
## 5 AL               2 fe         -1.51
## 6 AL               2 ml         -0.543
```

Let’s compare multilevel and fixed effects model fits



Let's compare multilevel and fixed effects model fits



Benefits of the multilevel approach: regularization

Our model *regularizes* (or shrinks) intercept and slope estimates by nudging them toward the global mean

This shift is larger when:

- The group mean (state here) is far away from the global (national) mean
- We have few observations for a unit (small N of counties here)

This shift is smaller when:

- The group mean is close to the global mean
- We have many observations for a unit (large N of counties)

Summary

What goes into a no-pooling (fixed effects) model

- Estimated intercepts are equal to empirical group means
- This approach typically *overfits* the data, and makes for poor prediction, unless we have a lot of data on each unit

The partial pooling alternative

- The complete pooling approach assumes all units are the same
- The no-pooling approach assumes all units are different
- The *partial pooling* approach assumes that all units are different, but come from the same larger population

Multilevel models are a good default

Multilevel models are very flexible and useful for modeling:

- Repeat observations of units over time
- Data with clusters

Multilevel models

- Shrink estimates toward group means
- Provide variance estimates across and within clusters
- Allow for flexible inference without overfitting