Intro to hierarchical models

Why multilevel regression modeling?

Consider a housing dataset that contains information about sales of 2000 houses across 100 different zipcodes.

```
housing_sales <- read_csv('http://math.montana.edu/ahoegh/teaching/stat408/datasets/HousingSales.csv')
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   City = col_character(),
##   State = col_character(),
##   Zip_Code = col_double(),
##   Living_Sq_Ft = col_double(),
##   Closing_Price = col_double()
## )
```

Q: Do you expect to see the same relationship between the size of the home `Living_Sq_Ft` and the sales price for all cities? Note a few cities in this dataset include Lincoln, NE; Lewistown, MT; Miami, FL; Honolulu, HI; Snowmass, CO; and Flint, MI.

So how should we model housing prices as a function of living space across these cities?

One option would be to use the indicator notation that we previously discussed:

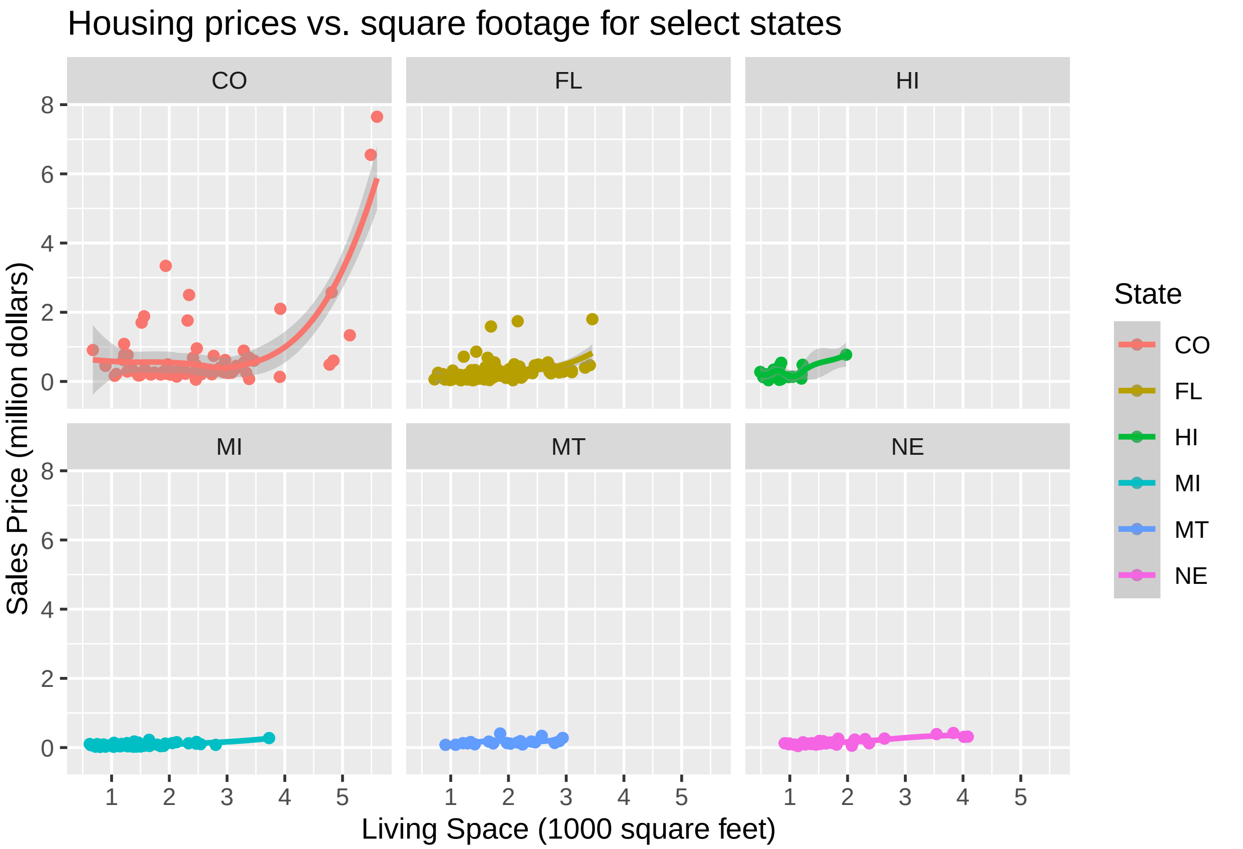
$$Y_i = \beta_0 + \beta_1 I_{i \in zip(1)} + ... + beta_{100} I_{i \in zip(100)} + \epsilon_i$$

where Y_i is the sales price of home i and $\epsilon_i \sim N(0, \sigma^2)$

Sketch out the mean housing price by zipcode as a function of living space using this model. Does this seem like a reasonable model?

```r
housing_sales %>% filter(State %in% c("NE", "MT", "CO", 'HI','FL','MI')) %>%
  mutate(sales_price = Closing_Price / 1000000, thousand_sq_ft = Living_Sq_Ft / 1000) %>%
  ggplot(aes(y = sales_price, x = thousand_sq_ft, color = State)) +
  geom_point() + geom_smooth(method = 'loess') +
  xlab('Living Space (1000 square feet)') +
  ylab('Sales Price (million dollars)') + facet_wrap(.~State) +
  ggtitle('Housing prices vs. square footage for select states')
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Another option would be to fit separate models for each zipcode. What are some of the implications for this type of model?

1. Different slopes and different intercepts

2. No information is shared across zipcodes

3. No mechanism for making predictions for zipcodes not in the sample (assuming this is permissable within the scope of inference)

A multilevel, or hierarchical model, contains another level that models the covariates from each individual level model.

Thus rather than

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

where $i = 1, ..., n$ corresponds to the n houses, the model can be written as

$$\begin{aligned} Y_i &= \alpha_{j[i]} + \beta_{j[i]} X_i + \epsilon_i && (1) \\ \alpha_j &= a_0 + b_0 u_j + \eta_{j1} && (2) \\ \beta_j &= a_1 + b_1 u_j + \eta_{j2} && (3) \end{aligned}$$

where $j[i]$ denotes the zipcode containing the i^{th} house. In this motivating dataset $i = 1, ..., 2000$ and $j = 1, ..., 100$. While X_i corresponds to house level covariates, u_j would be zipcode level covariates, and η are independent error terms.

Terminology

These multilevel or hierarchical models carry this designation for two reasons:

1. There are multiple levels in the data structure. In this case, consider houses nested in zipcodes.

2. The model also has multiple levels.

Multilevel models could also be applied for several layers...

About Mixed Models / Random Effects The authors intentionally avoid the term “random effects” and hence, mixed models. More on this later...

GH include several interesting applications of hierarchical models from their own research. Read through these in Chapter 1.2.

Motivations for using hierarchical models **Learn about treatment effects that may vary:** Some variables may have different impacts across groups, hierarchical models provides a formal way to address these questions.

Use all of the data to perform inferences for groups with small sample size: The hierarchical model allows the group level values to be informed by both the data in that group *and* the group level values from other groups.

Prediction: Hierarchical models provide a natural way for making predictions for new observations in an existing group and even new observations in a new group.

Analysis of Structured Data: The hierarchical structure provides a natural way to model data with inherent structure. Furthermore, there are natural extensions for data with to repeated measures, longitudinal, spatial, temporal, or spatiotemporal structure.

More efficient inference for regression parameters: This provides an alternative between separate models with no pooling and one model with complete pooling. Think of this as a data-driven partial pooling procedure.

Including predictors at multiple levels The model we have previously discussed for housing prices would permit using covariates for both the house-level and the zipcode-level, something that is difficult or impossible to do what standard model specifications.

Getting the right standard error accurately accounting for uncertainty in prediction and estimation: Consider cases where there is correlation across groups. The book touches on election results… If Indiana votes for Trump, does that make it more likely that Ohio will too? Multilevel models also give natural uncertainty estimates for new groups.

Multilevel Models

For multilevel models, observations fall into groups and coefficients can vary by the group.

Assume there are J groups and $j[i]$ denotes that observation i falls into group j

$$y_i = \alpha + \beta x_i + \epsilon_i$$

complete pooling

$$\begin{aligned}
y_{[1]i} &= \alpha_1 + \beta_1 x_{[1]i} + \epsilon_i \\
y_{[2]i} &= \alpha_2 + \beta_2 x_{[2]i} + \epsilon_i \\
&. \\
&. \\
&. \\
y_{[J]i} &= \alpha_J + \beta_J x_{[J]i} + \epsilon_i
\end{aligned}$$

no pooling

$$y_i = \alpha_{j[i]} + \beta_{j[i]} x_i + \epsilon_i$$

partial pooling

Shrinkage Equation Assume that the multilevel model only includes group averages, then the partial pooling estimate of the mean (or intercept) is

$$\hat{\alpha}_j \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}}$$

where σ_y^2 is the variance of the data and σ_α^2 is the variance of the group-level averages.

Thus the estimate value for a group is a weighted average from the data in that group and the overall data.

The weights are:

1. a function of the data variance and the number of observations in a group, and
2. function of the variance of the group level estimates.

For each scenario, large variance corresponds to a lower weight on that component and smaller variance (high precision) corresponds to higher weights

Correlation structure A common assumption in regression models is that the observations are independent. There are a few common data types that violate this assumption and can be addressed with hierarchical models.

Repeated Measurements: repeated measurements on persons (or units), thus the data observations are clustered.

Cross Sectional Data (Longitudinal/Time series): Repeated measurements across time.

"Fixed vs. Random" These type of models are commonly referred to as "mixed models" that include "fixed" and "random" effects.

Random Effects: the coefficients that vary (across groups) are often referred to as random effects. We will see a formal statistical distribution associated with these later on.

Fixed Effects: GH point out inconsistencies with this term. Fixed effects generally refer to coefficients that do not vary (say a parameter estimated across all groups). This could also apply to the separate models approach. The defining feature is largely a probability distribution for model.

Some general advice about when to use fixed/random effects focuses on the research goal; however, GH suggest *always* using multilevel models.

Furthermore, given the inconsistencies in the meaning of fixed/random, GH (and I) prefer using multilevel or hierarchical models.

Multilevel Modeling Pros and Cons

Classical Regression Overview

- prediction for continuous or discrete outcomes
- fitting of nonlinear relationships (using transformations and basis functions)
- inclusion of categorical predictors using indicator functions
- interactions between inputs
- GLM frameworks for non-Gaussian (normal) probability distributions

Multilevel Modeling

- Accounting for and estimating individual- and group-level variation by estimating group-level coefficients (and potentially including group-level covariates.
- Modeling variation among individual-level regression coefficients and making predictions for new individuals/groups.
- Note there is extra complexity in fitting a multilevel model and additional modeling assumptions.
- Limiting cases of multilevel models
  - very little group variation, then the multilevel model approaches the complete pooling scenario
  - very large group variation, then the multilevel model approaches the seperate model solution