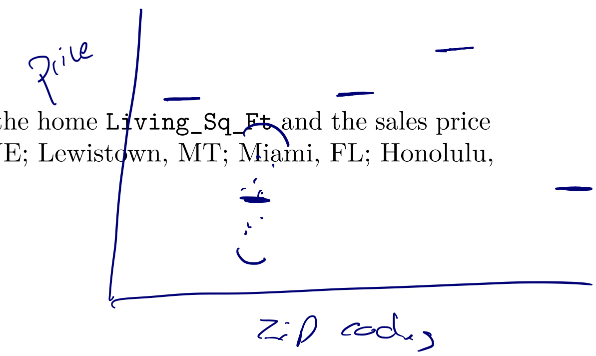
Lecture 4: Gelman Hill Ch 1

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')
```

```
## Parsed with column specification:
## cols(
##   City = col_character(),
##   State = col_character(),
##   Zip_Code = col_double(),
##   Living_Sq_Ft = col_double(),
##   Closing_Price = col_double()
## )
```



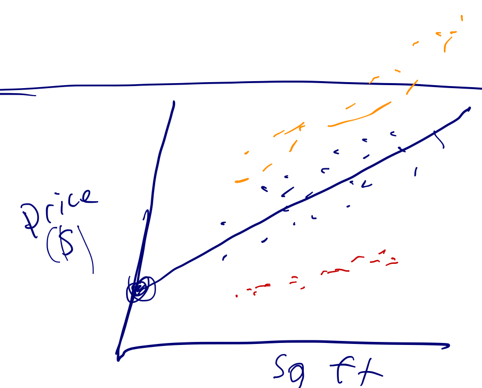
Q: Do you expect to see the same relationship between the size of the home `Living_Sq_Ft` and the sales price for all cities? Note a few cities in this dataset include Lincoln, NE; Lewistown, MT; Miami, FL; Honolulu, HI; Snowmass, CO; and Flint, MI.

One option is

$$Y_i = B_0 + B_1 I_{zip=(cat\,2),i} + B_{100} I_{zip=(cat\,101),i} + \varepsilon_i$$

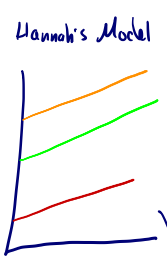
Y_i is the sales price

another option would be



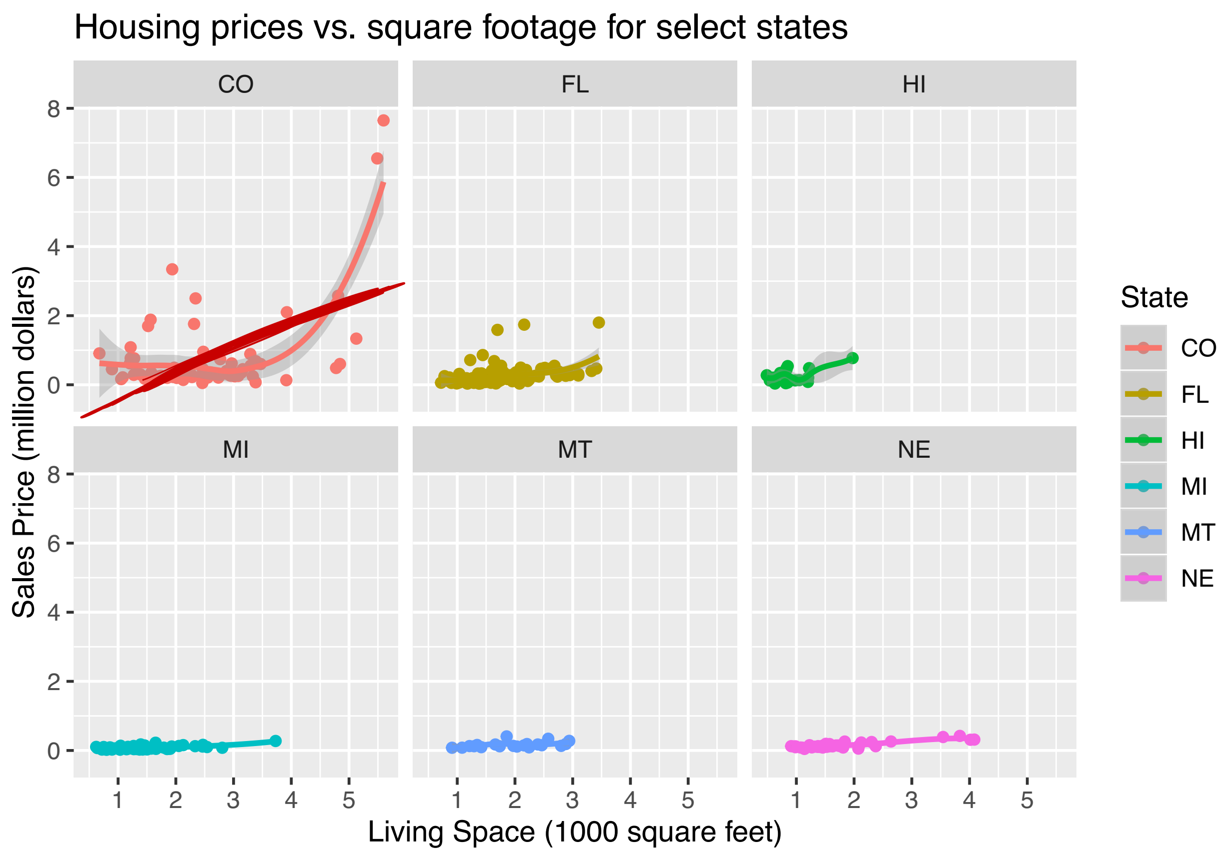
$$Y_i = B_0 + B_1 X_{sqft,i} + \varepsilon_i$$

$X_{sqft,i}$ is the square footage of house i



$$Y_i = B_1 I_{zip=cat(1),i} + \cdots B_{100} I_{zip=cat(100)} + \alpha_1 X_{sqft,i} I_{zip(cat(1)),i} + \alpha_{100} X_{sqft,i} I_{zip=cat(100)} + \varepsilon_i$$

```
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')
```



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

Q1. How is information shared across zipcodes?

Q2. How do we make predictions / inferences about houses not in the sample?

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 $\epsilon_i \sim N(0, \sigma_\epsilon^2)$

$$Y_i = \boxed{\alpha_{j[i]} + \beta_{j[i]} X_i} + \epsilon_i \quad \text{or} \quad Y_i \sim N(\alpha_{j[i]} + \beta_{j[i]} X_i, \sigma_\epsilon^2)$$

$$\alpha_j = \underline{a_0 + b_0 u_j} + \eta_{j1} \quad \text{or} \quad \alpha_j \sim N(a_0 + b_0 u_j, \sigma_{\eta_1}^2)$$

$$\beta_j = \underline{a_1 + b_1 u_j} + \eta_{j2} \quad \text{or} \quad \beta_j \sim N(a_1 + b_1 u_j, \sigma_{\eta_2}^2)$$

$u_j \sim$ group level covariates (school-district) $\quad \eta_{j1} \sim N(0, \sigma_{\eta_1}^2)$

$\eta_{j2} \sim N(0, \sigma_{\eta_2}^2)$

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 will have different impacts across groups.

Use all of the data to perform inferences for groups with small sample size:

The hierarchical model allows the group level variables to be informed by data in that group AND group level variables from other groups.

Prediction:

Hierarchical models provide a natural way for making predictions

* new observations in an existing group
* new observations in a new group

Analysis of Structured Data:

Hierarchical models are useful for structured data.

- See 534 spatial statistics
- See 536 time series

532 - Bayes

More efficient inference for regression parameters:

alternative between no pooling + complete pooling

Stein's Paradox

Including predictors at multiple levels

e.g. zipcode level info can be included in the model

Getting the right standard error accurately accounting for uncertainty in prediction and estimation:

* consider correlation across groups.

Features of GH and this course

GH “present methods and software that allow the reader to fit complicated, linear or nonlinear, nested or non-nested models. They emphasize the use of the statistical software packages R and BUGS. . . ” (Stan too)

“Most books define regression in terms of matrix operations. GH avoid much of this matrix algebra for the simple reason that it is now done automatically by computers. We are more interested in understanding the ‘forward’ or predictive, matrix multiplacation $X\beta$ than the more complicated inferential formula $(X^T X)^{-1} X^T y$. The latter computation and its generalizations are important but can be done out of sight of the user.”

GH "try as much as possible to display regression results graphically rather than through tables. . . GH consider graphical display of model estimates to be not just a useful teaching method but also a necessary tool in applied research.

Statistical texts commonly recommend graphical displays for model diagnostics. These can ve very useful,. . . but here we are emphasizing graphical displays of the fitted models themselves. It is our experience that, even when a model fits data well, we have difficulty understanding it if all we do is look at the tables of regresssion coefficients."

“Ultimately, you have to learn these methods by doing it yourself, and this chapter (CH1) is intented to make things easier by recounting stories about how we learned this by doing it ourselves. But we warn you ahead of time that we include more of our successes than our failures.”

Costs and benefits of GH approach

Using the GH approach to statistics is not easy. The challenge is not mathematical, but conceptual and computational.

Fitting classical regression and generalized regression models is easy in R, but graphing the results and simulating predictions can be a challenging programming expercise.

With multilevel modeling, model fitting is more complicated. (using R and Bugs and STAN)

The multilevel approach does have several advantages: - You can fit a greater variety of model - By working with simulations (rather than simply point estimates of parameters), you can directly capture inferential uncertainty and propogate it into predictions.

Course Structure

The course will initially focus on using R to

1. fit linear and generalized linear models,
2. graph data and estimated models, and
3. use simulation to propogate uncertainty in inferences and prediction.

Then the focus will shift to hierarchical models.