**Notes from Data Analysis Using Regression and Multilevel/Hierarchical Models - Andrew Gelman & Jennifer Hill**

Chapter 1: Why? (page 1 – 8)

* What is multilevel regression modeling?
  + Classical regression model:
  + Multilevel regression model:
    - * *Where the subscript ‘j’ index levels/groups.*
  + Generally, multilevel models are considered to be a regression (a linear or generalized linear model) in which the parameters – the regression coefficients – are given a probability model.
    - The second-level model has parameters of its own – the hyperparameters of the model – which are also estimated from data.
  + The two key parts of a multilevel model are:
    - Varying coefficients
    - A model for those varying coefficients (which can itself include group-level predictors)
  + While classical regression can sometimes accommodate varying coefficients by using indicator variables, multilevel models are set apart by the model of the variation between groups.
* Labels
  + Multilevel models are also called hierarchical, for two different reasons:
    - From the structure of the data (i.e. students clustered within schools)
    - From the model itself, which has its own hierarchy, with the parameters of the within-school regressions at the bottom, controlled by the hyperparameters of the upper-level model.
  + Multilevel models are often known as *random-effects* or *mixed-effects models*.
    - The regression coefficients that are being modeled are called *random effects,* in the sense that they are considered random outcomes of a process identified with the model that is predicting them.
    - By contrast, *fixed effects* correspond to either to parameters that do not vary (i.e., fitting the same regression line for each of the schools) or to parameters that vary but are not modeled themselves (i.e., fitting a least squares regression model with various predictors, including indicators for the schools).
      * Can be viewed as special cases of random effects, in which higher-level variance is set to 0 or ∞.
      * Hence, in the framework, all regression parameters are “random,” and the term “multilevel” is all encompassing.
  + A *mixed-effects* model includes both fixed and random effects.
* Motivations for multilevel modeling
  + Can be used for a variety of inferential goals including causal inference, prediction, and descriptive modeling.
    - Learning about treatment effects that vary
      * One of the basic goals of regression analysis is estimating treatment effects.
        + How does *y* change when some *x* is varied, with all other inputs held constant?
      * In many cases, it’s not the overall effect of *x* that is of interest, but how this effect varies in the population.
        + In classical statistics we study this variation using *interactions*.
      * Multilevel models also allow us to study effects that vary by group
        + i.e. an intervention that is more effective in some schools than others (perhaps because of unmeasured school-level factors such as teacher morale).
    - Using all the data to perform inferences for groups with small sample size
      * Classical estimation just using the local information can be essentially useless if the sample size is mall in the group.
      * Conversely, a classical regression ignoring group indicators can be misleading in group-level variation.
      * Multilevel modeling allows the estimation of group averages and group-level effects, compromising between the overly noisy within group estimate and the oversimplified regression estimate that ignores group indicators.
    - Prediction
      * Regression models are commonly used for predicting outcomes for new cases. But what if the data varies by group? Then we can make predictions for new units in existing groups or in new groups.
        + The latter is difficult to do in classical regression:

If a model ignores group effects, it will tend to understate the error in predictions for new groups.

But a classical regression that includes group effects does not have any automatic way of getting predictions for a new group.

* + - * A natural attack on the problem is a two-stage regression, first including group indicators and then fitting a regression of estimated group effects on group-level predictors.
        + Then, one can forecast for a new group, with the group effect predicted from the group-level model, and then the observations predicted from the unit-level model.
        + HOWEVER, if sample sizes are small in some groups, it can be difficult or even impossible to fit such a two-stage model classically, and fully accounting for the uncertainty at both levels leads directly to a multilevel model.
    - Analysis of structured data
      * Data that contains inherent multilevel structure (i.e. students within schools, patients within hospitals, data from cluster sampling) should include statistical methods (both sampling-theory or Bayesian) where inference includes the factors used in the design of data collection.
      * Multilevel modeling is a direct way to include indicators for clusters at all levels of a design without being overwhelmed with problems of overfitting from applying least squares or maximum likelihood to problems with large numbers of parameters.
    - Efficient inference for regression parameters
      * Traditional alternatives to multilevel modeling are *complete pooling*, where differences between groups are ignored, and *no pooling,* where data from different sources are analyzed separately.
      * *No pooling* ignores information and can give unacceptably variable inferences, and *complete pooling* suppresses variation that can be important or even the main goal of a study.
      * The extreme alternatives can in fact be useful as preliminary estimates, but ultimately *partial pooling* that comes out of a multilevel analysis is preferable.
    - Including predictors at two different levels
    - Getting the right standard error: accurately accounting for uncertainty
      * To get an accurate measure of predictive uncertainty, one must account for correlation of the outcome between states in a given election year.
        + Multilevel modeling is a convenient way to do this.
      * For certain types of predictions, multilevel models are required.
        + (i.e. a model of test scores for students within schools.)

Classical regression – school-level variability might be modeled by including an indicator variable for each school.

This framework makes it impossible to make a prediction for a new student in a new school, because there would not be an indicator for this new school in the model.

Chapter 11: Multilevel structures (page 237 - )

* Multilevel models are extensions of regression models where data are structured in groups and coefficients can vary by group.
* Varying-intercept and varying-slope models
  + With grouped data, a regression that includes indicators for groups is called a varying-intercept model because it can be interpreted as a model with a different intercept within each group.
  + Varying-intercept model:
  + Varying-slope model:
  + Varying-intercept, varying-slope model:
  + The varying slopes are interactions between the continuous predictor *x* and the group indicators.
* To make use of the multilevel structure of the data, we need to construct two data matrices, one for each level of the model.
  + Practically, the two-matrix format has the advantage that it contains each piece of information exactly once whereas a single large matrix can have each groups data repeated several times.
    - If group-level information needs to be added or changed, the single-matrix format invites errors.
    - Conceptually, the two matrix, or multilevel, data structure has the advantage of clearly showing which information is available on individuals and which on cities.
      * It also gives more flexibility in fitting models, allowing us to move beyond the classical regression framework.
* Individual- and group-level models
  + *Individual-level regression*
    - Classical regression notation: , where *X* includes the constant term, the treatment, and other predictors.
    - This individual-level regression has the problem that it ignores city-level variation beyond that explained by group-level predictors in the model.
  + *Group-level regression on city averages*
    - Has the advantage that its errors are automatically at the city level. However, by aggregating, it removes the ability of individual predictors to predict individual outcomes.
  + *Individual-level regression with city indicators, followed by group-level regression of the estimated city effects.*
    - Two steps:
      * Fitting a logistic regression to the individual data y given individual predictors along with indicators for the 20 cites.
      * Perform a linear regression at the city level, considering the estimated coefficients of the city indicators.
    - Reasonable but can run into problems when sample sizes are small in particular groups, or when there are interactions between individual- and group-level predictors.
    - Multilevel modeling is a more general approach that can include predictors at both levels at once.
* Multilevel models
  + A simple multilevel model would have two components: a logistic regression predicting the binary outcome given individual-level predictors and with an intercept that can vary by city, and a linear regression predicting the city intercepts from a city-level predictors.
* Repeated measurements, time-series cross sections, and other non-nested structures
  + Another kind of multilevel data structure involves repeated measurements on persons (or other units) – thus, measurements are clustered within persons, and predictors ca be available at the measurement or person level.
  + *Time-series cross-sectional data*
    - In settings where overall time trends are important, repeated measurement data are sometimes called time-series cross-sectional.
    - Time-series cross-sectional data are typically (although not necessarily) “rectangular” in structure, with observations at regular time intervals. In contrast, general repeated measurements could easily have irregular patterns (for example, in the smoking study, some children could be measured only once, others could be measured monthly and others yearly).
    - In addition, time-series cross-sectional data commonly have overall time patterns, for example, the steady expansion of the death penalty from the 1970s through the early 1990s.
  + Other non-nested structures
    - Non-nested data can also come from when individuals are characterized by overlapping categories of attributes.
* Indicator variables and fixed or random effects
  + *Classical regression: including a baseline and J – 1 indicator variables.*
    - When including an input variable with J categories into a classical regression, standard practice is to choose one of the categories as a baseline and include indicators for the other J – 1 categories.
  + *Multilevel regression: including all J indicators*
    - In a multilevel model it is unnecessary to do this arbitrary step of picking one of the levels as a baseline. In a classical regression these may no all be included because they would be collinear with the constant term, but in multilevel model this is not a problem because they are themselves modeled by a group-level distribution (which itself can be a regression)
  + *Fixed and random effects*
    - The varying coefficients in a multilevel model are sometimes called *random effects*, a term that refers to the randomness in the probability model for the group-level coefficients.
    - The term *fixed effects* is used in contrast to random effects - but not in a consistent way! Fixed effects are usually defined as varying coefficients that are not themselves modeled.
      * A classical regression including J – 1 = 19 city indicators as regression predictors is sometimes called a “fixed-effects model” or a model with “fixed effects for cities.” However, “fixed effects models” sometimes refer to regressions in which coefficients do not vary by group (so that they are fixed, not random).
    - When to use fixed effects (in the sense of varying coefficients are unmodeled) and when to use random effects?
      * Some say that fixed effects are appropriate if group-level coefficients are of interest, and random effects are appropriate if interest lies in the underlying population. Others recommend fixed effects when the groups in the data represent all possible groups, and random effects when the population includes groups not in the data. These two recommendations (and others) can be unhelpful.
      * This books advice, to always use multilevel modeling (“random effects”).
* Cost and benefits of multilevel modeling
  + Motivations for multilevel modeling
    - Accounting for individual- and group-level variation in estimating group-level regression coefficients.
    - Modeling variation among individual-level regression coefficients. In classical regression, one can do this using indicator variables, but multilevel modeling is convenient when we want to model the variation of these coefficients across groups, make predictions for new groups, or account for group-level variation in the uncertainty for individual-level coefficients.
    - Estimating regression coefficients for particular groups.
  + Complexity of multilevel models
    - Multilevel models have additional complexity of coefficients varying by group.
  + Additional modeling assumptions
    - Multilevel models requires additional assumptions beyond those of classical regression. Each level of the model corresponds to its own regression with its own set of assumptions such as additivity, linearity, independence, equal variance, and normality.
  + When does multilevel modeling make a difference?
    - Normal alternative to multilevel modeling is classical regression 0 either ignoring group-level variation, or with varying coefficients that are estimated classically (and not themselves modeled) – or combinations of classical regressions such as the individual and group-level models.
    - In some cases, both approaches coincide. When there is very little group-level variation, the multilevel model reduces to classical regression with no group indicators; conversely, when group-level coefficients vary greatly (compared to their standard errors of estimation), multilevel modeling reduces to classical regression with group indicators.
    - When the number of groups is small (less than five, say), there is typically not enough information to accurately estimate group-level variation. As a result, multilevel models in this setting typically gain little beyond classical varying-coefficient models.