36-463/663: Multilevel & Hierarchical Models

(P)review: in-class midterm

Brian Junker
132E Baker Hall
brian@stat.cmu.edu

In-class midterm

- Closed book, closed notes, closed electronics (otherwise I have to make it harder!)
- Mainly conceptual/theoretical material
  - No proofs, but maybe "calculation with greek letters"
- Some computer output to interpret
- Sketch some graphs
- Produce a couple of fitted values, residuals, confidence intervals, variances, by hand
  - Bring a calculator!
  - Don't use calculator on your phone/tablet

Generalities

- Sampling model vs measurement error model in statistics
- Regression models typically combine both sampling and measurement error (how?)
- Distributions as models
  - can represent the population from which the sample comes (sampling model)
  - can represent messy or unknown parts of data generation (meas error or generative model)

Basic Statistics

- Confidence intervals
  - *Traditional*: (Pt. Estimate) +/- 2 SE's
  - *Simulation*: Simulate values of estimated parameter (or simulate data and recalculate parameter estimate), cut off lower 2.5% and upper 97.5%
- Hypothesis testing
  - *Traditional*: H_0 induces a distribution on test statistic T. If T(real data) in the tail of this distribution, reject H_0
  - *Simulation*: simulate data from the fitted model under H_0 to get a distribution for T. If T(real data) in the tail of this distribution, reject H_0.

Regression I

- Interpreting slope
  - predictive or summary interpretation
  - causal or counterfactual interpretation
  - multiple predictors
- Interpreting intercept
  - transforming X's to make better interpretation
- Fit, diagnostics
  - does it make sense?
  - linearity, additivity of errors
  - "normal", iid residuals
  - R^2, F-test, AIC

Regression II

- Different ways of writing the model
  - lazy way $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
  - long way $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i,\ i = 1, \ldots, n$
  - matrix way $Y = X\beta + \varepsilon$
- normal distribution, variances, covariances, means
- estimating $E[y_{new}]$ vs predicting y_{new}
- Using simulation to display uncertainty
  - simulating new data
  - simulating new regression coefficients

Regression III

- Transformations
  - transforming y can help normality/symmetry
  - transforming x can help with influential data
- Model building (for predictive modeling)
  - include a term if it has the ‘right’ sign
  - exclude a term if it has ‘wrong’ sign and not significant
  - think hard if term has ‘wrong’ sign and is significant
- Assessing fit
  - resid plots
  - R^2
  - F-tests
  - AIC/BIC (difference of 2 interesting, difference of 3 a big deal)

Generalized Linear Models (GLMs)

- Generalized linear model
  - $y_i \sim f(y|\mu_i, \ldots)$, $\mu_i = E[y_i]$
  - $g(\mu_i) = \theta_i = X_i \beta$
- Examples:
  - Normal linear regression
  - Logistic regression
  - Poisson regression
  - probit regression
  - multinomial logit regression

Logistic Regression

- Shape of $g^{-1}(X\beta)$
- $\log(p/1-p) = X\beta$; Interpreting coefficients
  - divide by 4 rule
  - logits (log odds)
  - log odds-ratio (difference in log odds)
- Assessing fit
  - resid plots (binned, not raw!)
  - deviance (chi-squared) tests
  - AIC/BIC (difference of 2 interesting, difference of 3 a big deal)
- Overfitting if fitted prob's close to 0 or 1

Other GLMs have similar (but different) interpretations…

- E.g. Poisson regression:
  - Shape of $g^{-1}(X\beta)$
  - $\log(\lambda) = X\beta$…
    - Interpreting coefficients
  - Assessing fit
    - resid plots (binned, not raw!)
    - deviance (chi-squared) tests
    - AIC/BIC (diff. of 2 interesting, difference of 3 a big deal)
  - AIC Deviations from the "standard" model
    - Overdispersion
    - Zero inflation, etc.

Causal Inference I

- Fundamental problem of causal inference
- ACE
- Unbiased estimate of β_1 = effect of T:
  - y = $\beta_0 + \beta_1$ T + ϵ: confounders independent of T
  - y = $\beta_0 + \beta_1$ T + β_2 X_2 + … + β_k X_k + ϵ: all possible confounders
- Include pre-treatment covariates (why?)
- Exclude post-treatment covariates (why?)
- Interpret obs study by analogy/comparison with randomized trail

Causal Inference II

- ***Causal inferences*** from observational studies:
  - *Instrumental variables* – substitute for the coin flip in randomized trials, to eliminate selection effects
  - *Propensity score matching* – rearrange the data to eliminate selection effects
  - *Regression discontinuity designs* – exploit random errors in selection effects
  - *Bounding the influence of confounders* – sometimes the effect (ACE) of T_i is so big, that we can calculate that no reasonable set of confounders could be responsible for it. *(This is basically how the link between smoking and lung cancer was made.)*
- ***Require strong assumptions about data generation!***

Multi-Level Models (MLMs)

- Compromise between full pooling and no pooling (partial pooling)
- replace two linear models "on top of each other" with a single model with two equations
- shrinkage in mlm estimates
- fixed effects, random effects, residuals

MLMs

- Hierarchical Model

  Level 2: $\alpha_j \overset{iid}{\sim} N(\beta_0, \tau^2)$

  Level 1: $y_i \overset{indep}{\sim} N(\alpha_{j[i]}, \sigma^2)$
- Multi-Level Model

$$\begin{aligned} y_i &= \alpha_{j[i]} + \epsilon_i, \ \epsilon_i \overset{iid}{\sim} N(0, \sigma^2) \\ \alpha_j &= \beta_0 + \eta_j, \ \eta_j \overset{iid}{\sim} N(0, \tau^2) \end{aligned}$$

- Variance-Components Model

$$y_i = \beta_0 + \eta_{j[i]} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$
$$\eta_j \overset{iid}{\sim} N(0, \tau^2)$$

- lmer() notation

  y ~ 1 + (1|group)

MLM Residual Analysis

- $Y = X\beta + Z\eta + \varepsilon$ (Laird-Ware Form)
- *Marginal residuals:*
  - Should be mean 0, but may show grouping structure
  - May not be homoskedastic! $Y - X\beta \quad (= Z\eta + \varepsilon)$
  - Good for checking fixed effects, just like linear regr.
- *Conditional residuals:* $Y - X\beta - Z\eta \quad (= \varepsilon)$
  - Should be mean zero with no grouping structure
  - Should be homoskedastic!
  - Good for checking normality of ϵ, outliers
- *Random effects:* $Y - X\beta - \varepsilon \quad (= Z\eta)$
  - Should be mean zero, but may show grouping structure
  - May not be be homoskedastic!
  - Good for checking normality of ε , outliers

All models

- Interpreting R output
- Calculating $E[y_i]$ or a predicted value from a fitted model
- Understanding residual variance, overdispersion
- Understanding variance components, correlation between units in same group/cluster
- Interpreting plots, e.g.
  - residuals and residual plots
  - fitted regression plotted on top of raw data
  - histograms, boxplots, qqplots, etc.

Summary

- Closed book, closed notes, closed electronics (otherwise I have to make it harder!)
- Mainly conceptual/theoretical material
  - No proofs, but maybe "calculation with greek letters"
- Some computer output to interpret
- Sketch some graphs
- Produce a couple of fitted values, residuals, confidence intervals, variances, by hand
  - Bring a calculator!
  - Don't use calculator on your phone/tablet