Multilevel Models

Bret Larget

Departments of Botany and of Statistics
University of Wisconsin—Madison

April 22, 2008

Reminder about Corn Example

- We consider a subset of a larger data set on corn grown on the island Antigua.
- The response variable we consider is the harvest weight (`harvwt`) per plot (units unknown).
- There are eight sites with eight separate plots within each site where the corn is grown under the same treatment conditions.

Multilevel Model

- In a multilevel model, we may have

$$y_i = \alpha_{j[i]} + e_i$$

where $i = 1, \ldots, 64$ indexes the observation and $j[i] = 1, \ldots, 8$ indicates which of the eight sites contains the ith observation.

  - $\alpha_j, \sim N(\mu_\alpha, \sigma_\alpha^2)$ are *modeled*;
  - $e_i \sim \text{iid } N(0, \sigma^2)$;
  - μ_α, σ_α, and σ^2 are unmodeled.

Likelihood

- We can express the likelihood of the data in terms of the unmodeled and modeled parameters.
- The normal density $f(x \mid \mu, \sigma)$ for mean μ and standard deviation σ is large when the x is close to μ.
- The likelihood has the form

$$L = \left(\prod_{j=1}^{J} f(\alpha_j \mid \mu_\alpha, \sigma_\alpha) \right) \left(\prod_{i=1}^{n} f(y_i \mid \alpha_{j[i]}, \sigma) \right)$$

- The left factor is large when all of the α_j are close to μ_α.
- The right factor is large when each α_j is close to the y_i in group j.
- The best estimate of μ_α will be the overall mean $\bar{y}_{\text{all}}$.
- The likelihood is maximized at a balancing point when the α_j is somewhere between the overall mean $\bar{y}_{\text{all}}$ and the sample means $\bar{y}_j$.

Pooling

- *Complete pooling* is the estimate that assumes no difference between groups so that
$$\alpha_1 = \alpha_2 = \cdots = \alpha_J = \bar{y}_{\text{all}}$$
This corresponds to an extreme multilevel model where $\sigma_\alpha = 0$.
- *No pooling* estimates each α_j using only data from the jth sample.
$$\alpha_1 = \bar{y}_j, \ \ldots, \ \alpha_J = \bar{y}_J$$
This corresponds to an extreme multilevel model where $\sigma_\alpha = +\infty$.
- Multilevel models correspond to *partial pooling* where data from other samples effects estimates for sample j, but the data within sample j is most influential.
- The estimated coefficients are a weighted average between the corresponding sample mean and the grad mean.
$$\hat{\alpha}_j \approx \frac{\left(\frac{n_j}{\sigma^2}\right)\bar{y}_j + \left(\frac{1}{\sigma_\alpha^2}\right)\bar{y}_{\text{all}}}{\frac{n_j}{\sigma^2} + \frac{1}{\sigma_\alpha^2}}$$

Weighted Average

$$\hat{\alpha}_j \approx \frac{\left(\frac{n_j}{\sigma^2}\right)\bar{y}_j + \left(\frac{1}{\sigma_\alpha^2}\right)\bar{y}_{\text{all}}}{\frac{n_j}{\sigma^2} + \frac{1}{\sigma_\alpha^2}} = \left(\frac{\frac{n_j}{\sigma^2}}{\frac{n_j}{\sigma^2} + \frac{1}{\sigma_\alpha^2}}\right)\bar{y}_j + \left(\frac{\frac{1}{\sigma_\alpha^2}}{\frac{n_j}{\sigma^2} + \frac{1}{\sigma_\alpha^2}}\right)\bar{y}_{\text{all}}$$

- An expression of the form
$$w_1 A + w_2 B$$
where $w_1 + w_2 = 1$ is a *weighted average* of A and B.
- The relative distance of the average to the endpoints is inversely proportional to the weights.
- (If one weight is ten times the other, the distance of the average to the end with the higher weight will be ten times smaller than to the other.)

Interpretations

$$\hat{\alpha}_j \approx \frac{\left(\frac{n_j}{\sigma^2}\right)\bar{y}_j + \left(\frac{1}{\sigma_\alpha^2}\right)\bar{y}_{\text{all}}}{\frac{n_j}{\sigma^2} + \frac{1}{\sigma_\alpha^2}}$$

- Here, when:
  - n_j gets bigger (more direct data in the group), or;
  - σ^2 gets smaller (more likely for y_i to be close to group means)

  $\hat{\alpha}_j$ moves closer to $\bar{y}_j$.
- When
  - σ_α^2 gets smaller (more likely for all α_j to be close to each other),

  $\hat{\alpha}_j$ moves closer to $\bar{y}_{\text{all}}$.

Numerical Example

- $n_j = 8$ for $j = 1, \ldots, 8$
- $\hat{\mu}_\alpha = \bar{y}_{\text{all}} = 4.29$
- $\hat{\sigma}_\alpha = 1.55$
- $\hat{\sigma} = 0.87$
- $\bar{y}_1 = 4.88$
- $\hat{\alpha}_1 = 4.86$

Site DBAN

- $\frac{8}{(0.87)^2} = 10.57$
- $\frac{1}{(1.55)^2} = 0.42$
- $$\frac{10.57}{10.99}(4.88) + \frac{0.42}{10.99}(4.29) = 4.86$$
- The estimate is shrunk very little toward the grand mean.

Site NSAN

- $\bar{y}_3 = 2.09$
- $\bar{y}_{\text{all}} = 4.29$
- $$\frac{10.57}{10.99}(2.09) + \frac{0.42}{10.99}(4.29) = 2.17$$
- The estimate is shrunk a little more toward the grand mean than for the other site since the same proportion of a longer distance is larger.