1 Statistics 249: Categorical Data/GLM Lecture 8

1.1 Mixed-Effect Models

1.1.1 What Mixed-Effects Models did we look at previously?

- Chick Weights
- Rat Lengths - (Biological Mothers vs. Nurturing Mothers)
- Catepillar Lengths - Growth rates between 49 different families
- Pituitary Length in children, a dentistry study

Consider the children are independent from each other, but their personal measurements, not exactly.

Main goal of Mixed-Effects is to fit a study on a small amount of data with potentially too many parameters. A random effect takes one degree of freedom for all of the many degrees of freedom certain fixed factors would take.

1.1.2 Toy example - Pure Random Effect

\[ Y_{ik} = \mu + T_i + \varepsilon_{ik} \]

\[ T_i \sim N(0, \sigma^2_T) \text{ independent} \]

\[ \varepsilon_{ik} \sim i.i.d N(0, \sigma^2) \]

\[ \varepsilon \perp T \]

1.1.3 example: response = \( f(\text{dosage}) + \text{error} \)

In this study, the order in which the doses are applied is considered independent of effect. However, does the linear relationship from lower to higher dosage need to be preserved, if we have an idea of the function.

In the beginning, there were parameteric forms for \( f \), linear, exponential, and other common parametric models. However, if the relationship seems unreasonably, a function

The use of the random effects is to make up for the structure we are not capable of adding to the mean, we choose instead to add to the error.
### 1.1 Mixed-Effect Models

#### 1.1.4 Balanced Design

- \( \text{SST} = \text{SST}_{\text{reat}} + \text{SSE} \)

- \( \mathbb{E} [\text{MSE}] = \sigma^2 \)

\[
\mathbb{E} [\text{SSE}] = \sum_{i,j} \mathbb{E} \left[ (Y_{ij} - \bar{Y}_.)^2 \right] = \sum_{i,j} \mathbb{E} \left[ \left( \frac{n-1}{n} \varepsilon_{ij} - \frac{1}{n} \sum_{k \neq j} \varepsilon_{ij} \right)^2 \right] = \sum_{ij} \left( \frac{n-1}{n} \right)^2 \sigma^2 + \frac{1}{n^2} (n-1)\sigma^2 = \sum_{ij} \frac{n-1}{n} \sigma^2
\]

#### 1.1.5 Finding the Mean squared of the treatment

\( \mathbb{E} [\text{MST}_{\text{reat}}] = c\sigma_T^2 + \sigma^2 \)

The formula being:

\[
\text{MST}_{\text{reat}} = \sum_{i,j} \left( \bar{Y}_{i.} - \bar{Y}_. \right)^2 = n \sum_i \left( \frac{1}{n} \sum_k \mu + T_i + \varepsilon_{ik} + \frac{1}{I} \sum_j T_j + \frac{1}{I} \sum_{i,k} \varepsilon_{ik} \right)^2
\]

\[
\bar{Y}_{i.} - \bar{Y}_. = \left( \frac{I-1}{I} T_i + \frac{1}{I} \sum_{k \neq i} T_k + \frac{1}{I} \sum_j T_j + \frac{1}{I} \sum_{i,k} \varepsilon_{ik} \right)^2
\]

At this point all of the terms are independent. If \( \mathbb{E} [T_i^2] = \sigma_T^2 \) and \( \mathbb{E} [\varepsilon_{ik}] = \sigma^2 \) we get:

\[
\left( \frac{I-1}{I} \right)^2 \cdot \sigma_T^2 + \left( \frac{1}{I} \right)^2 \cdot (I-1)\sigma_T^2 + \frac{(I-1)}{I^2n^2} \cdot n\sigma^2 + \frac{nI-n}{I^2n^2} \sigma^2
\]

Summing the \( I \) terms of these and then dividing by \( I^2 \) results in the answer \( \sigma_T^2 + \frac{1}{n} \sigma^2 \). In general, the mean squared of treatment is an unbiased estimate of \( \sigma^2 \) (It becomes a non-central chi-squared distribution), however, in the null hypothesis \( \sigma_T^2 \) is 0, so the MST is a good estimate of the Error.
1.1.6 Hypothesis Testing

In the fixed effect model we used \( \frac{\text{MST}_{\text{real}}}{\text{MSE}} \). But now the hypothesis is:

\[
\begin{align*}
H_0 : \sigma^2_T &= 0 \\
H_A : \sigma^2_T &> 0
\end{align*}
\]

Under the null model \( \frac{\text{MST}_{\text{real}}}{\text{MSE}} \sim F_{I-1,N-I} \).

Choosing a critical value \( F_{I-1,N-I,\alpha} \) then if the ratio is greater than the critical value, one should reject the null hypothesis \( H_0 \) at a \( \alpha \) level of significance.

1.1.7 Under the null model

- \( \frac{\text{SST}_{\text{real}}}{\sigma^2} \sim \chi^2_{I-1} \)
- \( \frac{\text{SSE}}{\sigma^2} \sim \chi^2_{N-I} \)
- \( \text{SST}_{\text{real}} \perp \text{SSE} \)

1.2 A big General Model

\[
\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \mathbf{\beta}_{p \times 1} + \mathbf{Z}_{n \times q} \mathbf{u}_{q \times 1} + \mathbf{\varepsilon}_{n \times 1}
\]

\[
\text{Var} \left( \begin{bmatrix} \mathbf{u} \\ \mathbf{\varepsilon} \end{bmatrix} \right) = \sigma^2 \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}_{(q+n) \times (q+n)}
\]

Consider \( \mathbf{u} \) a random vector particular to our individuals. With the \( R \) matrix, one can introduce heteroskedacity of the error (different variances).

The maximum likelihood estimate of the variance, without knowing the mean, are going to be biased.

\[
f \left( \mathbf{Y}; \beta_0, \beta_1, \sigma^2 \right) = \left( \frac{1}{2 \pi \sigma^2} \right)^{n/2} \exp \left( -\frac{1}{2} \sigma^2 \sum \left( \frac{y_i - \beta_0 - \beta_1 x_1}{\sigma} \right)^2 \right)
\]

And so:

\[
\ell \left( \mathbf{Y}; \beta_0, \beta_1, \sigma^2 \right) = -\frac{n}{2} \log(2\pi \sigma^2) - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_1)^2
\]

\[
\frac{\partial}{\partial \beta_0} \ell \left( \mathbf{Y}; \beta_0, \beta_1, \sigma^2 \right) = 0 \Rightarrow \sum (y_i - \beta_0 - \beta_1 x_1 = 0)
\]

\[
\frac{\partial}{\partial \beta_1} \ell \left( \mathbf{Y}; \beta_0, \beta_1, \sigma^2 \right) = 0 \Rightarrow \sum x_1 (y_i - \beta_0 - \beta_1 x_1 = 0)
\]

Estimates become:

- \( \beta_0 = \frac{\sum y_i - \beta_1 \sum x_i}{n} \)
1.3 Maximum Likelihood method or Restricted Method

- $\beta_1 = \frac{\sum x_i(y_i - \bar{y})}{\sum x_i + \sum x_i^2}$

Let's say we do not know $\sigma^2$:

\[
\frac{\partial}{\partial \sigma^2} \ell \Rightarrow -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = 0
\]

And the estimate for $\sigma^2$ that is mle $\hat{\sigma}^2 = \frac{\text{SSE}}{n}$. But using the Restricted Maximum likelihood method, ReML, we get the result $\hat{\sigma}^2_{\text{REML}} = \frac{\text{SSE}}{n-1}$.

1.3 Maximum Likelihood method or Restricted Method

For the more complicated General Linear model, we are going to look for a restricted likelihood. The “Justification” for REML is that the equation works for the simple model, so it might as well work for the general model.

\[
f(\tilde{Y}|u) \sim N_n(\tilde{X}\hat{\beta} + Zu, \sigma^2 R)
\]

The formula for Multivariate normal:

\[
f(\tilde{Y}|u) = |R|^{-1/2} \left( \frac{1}{2\pi \sigma} \right)^{n/2} \exp \left( -\frac{1}{2\sigma^2} \left( \tilde{Y} - \tilde{\beta} - Zu \right)^T R^{-1} \left( \tilde{Y} - \tilde{\beta} - Zu \right) \right)
\]

\[\bullet u \sim N(0, \sigma^2 G)\]

\[\bullet f(u) = |G|^{-1/2} \left( \frac{1}{2\pi \sigma} \right)^{q/2} \exp \left( -\frac{1}{2\sigma^2} U' G^{-1} U \right)\]

To remove $u$ we receive:

\[
f(Y; u) = f(u) \cdot f(u|u) = (2\pi \sigma^2)^{(n+q)/2} |R|^{-1/2} |G|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \left( \tilde{Y} - \tilde{\beta} - Zu \right)^T R^{-1} \left( \tilde{Y} - \tilde{\beta} - Zu \right) + U' G^{-1} U \right] \right\}
\]

We choose to minimize with respect to $\tilde{\beta}$ and $u$.

Call

\[\left( \tilde{Y} - \tilde{\beta} - Zu \right)^T R^{-1} \left( \tilde{Y} - \tilde{\beta} - Zu \right) \text{ minimized through weighted least squares.}\]

\[U' G^{-1} U \text{ the random effect.}\]

Sets:

\[\bullet X^T R^{-1} \left( \tilde{Y} - X\hat{\beta} - Z - u \right) = 0\]

\[\bullet \text{And } Z^T R^{-1} \left( \tilde{Y} - X\hat{\beta} - Z - u \right) + G^{-1} \hat{U} = 0\]

\[\bullet X^T R^{-1} X\hat{\beta} = (-Z\hat{u} + X^T R^{-1} Y)\]

\[\hat{\beta} = (X^T R^{-1} X)^{-1} X^T R^{-1} Y - Z\hat{u}\]