Statistics 203: Introduction to Regression and Analysis of Variance

*Mixed Effects*

Jonathan Taylor
Today’s class

- Mixed effects two-way ANOVA – what we didn’t cover last class.
- Mixed effects models – random intercepts.
- General form of a mixed effect model.
- Details: MLE and REML.
In some studies, some factors can be thought of as fixed, others random.

For instance, we might have a study of the effect of a standard part of the brewing process on sodium levels in the beer example.

Then, we might think of a model in which we have a fixed effect for “brewing technique” and a random effect for beer.
Two-way mixed effects model

\[ Y_{ijk} \sim \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}, \quad 1 \leq i \leq r, 1 \leq j \leq m, 1 \leq k \leq n \]

\[ \varepsilon_{ijk} \sim N(0, \sigma^2), \quad 1 \leq i \leq r, 1 \leq j \leq m, 1 \leq k \leq n \]

\[ \alpha_i \sim N(0, \sigma^2_\alpha), \quad 1 \leq i \leq r. \]

\[ \beta_j, \quad 1 \leq j \leq m \] are constants.

\[ (\alpha\beta)_{ij} \sim N(0, (m - 1)\sigma^2_{\alpha\beta}/m), \quad 1 \leq j \leq m, 1 \leq i \leq r. \]

Constraints:
- \[ \sum_{j=1}^{m} \beta_j = 0 \]
- \[ \sum_{i=1}^{r} (\alpha\beta)_{ij} = 0, \quad 1 \leq i \leq r. \]
- \[ \text{Cov} ((\alpha\beta)_{ij}, (\alpha\beta)_{i'j'}) = -\sigma^2_{\alpha\beta}/m \]

\[ \text{Cov}(Y_{ijk}, Y_{i'j'k'}) = \]
\[ \delta_{jj'} \left( \sigma^2_{\beta} + \delta_{ii'} \frac{m-1}{m} \sigma^2_{\alpha\beta} - (1 - \delta_{ii'}) \frac{1}{m} \sigma^2_{\alpha\beta} + \delta_{ii'} \delta_{kk'} \sigma^2 \right) \]
ANOVA tables: Two-way (mixed)

<table>
<thead>
<tr>
<th>SS</th>
<th>df</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSA</td>
<td>r – 1</td>
<td>$\sigma^2 + nm\sigma^2_\alpha$</td>
</tr>
<tr>
<td>SSB</td>
<td>m – 1</td>
<td>$\sigma^2 + nr\sum_{i=1}^{m} \beta_i^2 + n\sigma^2_{\alpha\beta}$</td>
</tr>
<tr>
<td>SSAB</td>
<td>(m – 1)(r – 1)</td>
<td>$\sigma^2 + n\sigma^2_{\alpha\beta}$</td>
</tr>
<tr>
<td>SSE</td>
<td>(n – 1)rm</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

- To test $H_0 : \sigma^2_\alpha = 0$ use $SSA$ and $SSE$.
- To test $H_0 : \beta_1 = \cdots = \beta_m = 0$ use $SSB$ and $SSAB$.
- To test $H_0 : \sigma^2_{\alpha\beta} \beta$ use $SSAB$ and $SSE$. 
Mixed linear models

- Not every model is an ANOVA!
- Suppose we study the effect of a blood pressure meant to lower blood pressure over time and we study \( r \) patients.
- For each patient we record BP at regular intervals over a week (every day, say).
- Drug will have varying efficacy in the population.
- Model

\[
Y_{ij} = \beta_0 + \gamma_i + \beta_1 X_{ij} + \varepsilon_{ij}
\]

- \( \varepsilon_{ij} \sim N(0, \sigma^2) \) i.i.d.
- \( \gamma_i \sim N(0, \sigma^2_\gamma) \) i.i.d. and independent of \( \varepsilon_{ij} \)'s
- \( X_{ij} = j, 1 \leq j \leq 7 \) in this example...
Random intercept model – balanced

- Model is called a “random intercept” model.
- It is “balanced” because every subject had the exact same $X$’s.

\[
\mathbb{E}(Y_{ij}) = \beta_0 + \beta_1 X_j.
\]

\[
\text{Cov}(Y) = \sigma^2 I_{nr \times nr} + \sigma^2 \gamma J_n = I_r \otimes V_0
\]

where $J_n$ is an $n \times n$ matrix with all 1’s;

\[
V_0 = \sigma^2 I_n + \sigma^2 \gamma J_n.
\]
A useful bit of notation for multivariate statistics.

Given two matrices $A_{m \times n}$ and $B_{p \times q}$, the Kronecker product $A \otimes B$ is an $mp \times nq$ matrix with entries

$$(A \otimes B)_{(i-1)p+l,(j-1)q+k} = A_{ij} \cdot B_{lk}.$$ 

Properties:

$$(A \otimes B)^t = A^t \otimes B^t$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

$$(A \otimes B)(X \otimes Y) = (AX) \otimes (BY)$$

$\text{rank}(A \otimes B) = \text{rank}(A) \text{rank}(B)$$

$\text{Tr}(A \otimes B) = \text{Tr}(A) \text{Tr}(B)$$

$\det(A_{a \times a} \otimes B_{b \times b}) = \det(A)^b \det B^a$
Generalized least squares

In a fixed effects model:

\[ Y = X\beta + \varepsilon \]

- We know what OLS estimators of \( \beta \) are:
  \[ \hat{\beta}_{OLS} = (X^tX)^{-1}X^tY \]
  if \( \varepsilon \sim N(0, \sigma^2) \) then they are MLE.
- If \( \varepsilon \sim N(0, D) \), \( D \) diagonal
  \[ \hat{\beta}_{WLS} = (X^tD^{-1}X)^{-1}X^tD^{-1}Y \]
  are MLE estimators.
- What if \( \varepsilon \sim N(0, V) \), \( V \) not diagonal? Then
  \[ \hat{\beta}_{GLS} = (X^tV^{-1}X)^{-1}X^tV^{-1}Y \]
  are MLE estimators. Proof is similar to OLS and WLS cases.
Fitting the random intercept model

- Likelihood has parameters $\beta_0, \beta_1, \sigma^2, \sigma^2_\gamma$.
- For $\sigma^2, \sigma^2_\gamma$ fixed

$$\hat{\beta}=(X^tV^{-1}X)(X^tV^{-1}Y), \quad V = I_r \otimes (\sigma^2 I_n + \sigma^2_\gamma J_n)$$

- ML estimates of $\beta_0, \beta_1$ turn out to be identical to fixed effects analysis

$$\hat{\beta}_1 = \frac{\sum_{j=1}^n X_j \bar{Y}.j - n \bar{X} \cdot \bar{Y}.}{\sum_{j=1}^n X_j^2 - n \bar{X}^2}$$

$$\hat{\beta}_0 = \bar{Y}. - \hat{\beta}_1 \bar{X}$$

where

$$\bar{Y}.j = \frac{1}{r} \sum_{i=1}^r Y_{ij}.$$
Variance-covariance of $\hat{\beta}$

- What about variance? In the generalized least squares setting

$$\text{Var}(\hat{\beta}) = (X^t V^{-1} X)^{-1}.$$ 

- In the balanced random intercept model

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2 + \sigma^2 / n}{r} + \frac{\sigma^2 \bar{X}^2}{r \sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{r \sum_{j=1}^{n} (X_j - \bar{X})^2}$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \bar{X} \cdot \text{Var}(\hat{\beta}_1).$$

- Similar to fixed effects: only difference is $\sigma^2_\gamma$ in $\text{Var}(\beta_0)$. 
Estimating $\sigma^2, \sigma^2_\gamma$

- If you go through the calculus, ignoring constraints $\sigma^2, \sigma^2_\gamma \geq 0$ the maximum likelihood estimator of $\sigma^2$ (denotes a “pre”-MLE) is

$$\hat{\sigma}^2 = \frac{\sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2}{r(n-1)} = \frac{SSE}{r(n-1)}$$

and

$$\hat{\sigma}^2_\gamma = \frac{1}{n} \left( \frac{SSA}{m} - \hat{\sigma}^2 \right)$$

where

$$SSE = \sum_{i,j} (Y_{ij} - \hat{Y}_{ij})^2, \quad SSA = \sum_{i,j} (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

- If $\hat{\sigma}^2_\gamma < 0$, then $\hat{\sigma}^2_\gamma = 0$ and

$$\hat{\sigma}^2 = \frac{SSA + SSE}{rn}.$$
Unbalanced data

- If not all the \(X\)’s are the same for each subject, or some observations are missing, things are more complicated.
- Likelihood has the same parameters only the variance covariance matrix is no longer a Kronecker product and MLE estimates are not closed form anymore.
- This is what computers are for!
- Two different techniques: MLE and REML.
Residual maximum likelihood

- Suppose that we have an $n - p \times n$ matrix $K$ of rank $n - p$ such that $K'X = 0$.
- Then

$$K'Y = K'(X\beta + \varepsilon) = K\varepsilon \sim N(0, K'VK).$$

- The distribution of $K'Y$ doesn’t have any $\beta$’s in it.
- We can use $K'Y$ to estimate parameters in $V (\sigma^2, \sigma_\gamma^2)$. This technique is called REML – residual maximum likelihood. Also works if $K$ is $n \times n$, only then the inverses have to be thought of as generalized inverses.
- Once $V$ has been estimated by $\hat{V}_{REML}$ we estimate

$$\hat{\beta}_{REML} = (X^t\hat{V}_{REML}^{-1}X)^{-1}X^t\hat{V}_{REML}^{-1}Y.$$ 

- We can approximate variance-covariance matrix as

$$\hat{\beta}_{REML} \sim (X^t\hat{V}_{REML}^{-1}X)^{-1}.$$
The random intercept model can be generalized quite a bit: random slopes, more than one predictor, etc.

The random/mixed ANOVA models and random intercept model all have the form

\[ Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + Z_{n \times q} \gamma_{q \times 1} + \epsilon_{n \times 1} \]

where

- \( \epsilon \sim N(0, \sigma^2 I) \);
- \( \gamma \sim N(0, D) \) for some covariance \( D \): simplest model, \( D \) is diagonal

In this model

\[ Y \sim N(X \beta, ZDZ' + \sigma^2 I). \]

Parameters to estimate: \( \beta \) and any parameters in \( D \). If \( D \) is diagonal, then there are \( q \) “variance” components to estimate.