Random Effects

Henry Scharf

MATH/STAT 571B

Module Goals:

Random Effects

Students will be able to:

1. Explain the conceptual difference between a fixed and random effect in a statistical model.
2. Implement foundational mixed effects models in R using the `lme4` package.

Random Effects

Random vs. Fixed Effects

$$y_{ij} = \mu + \underbrace{\tau_i}_{\text{fixed}} + \underbrace{\beta_j}_{\text{random}} + \epsilon_{ij}, \quad \beta_j \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\beta^2), \quad \epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$$

- When the *level* of an effect can be set, then generally the treatment should be treated as fixed (what we've done so far).
- When the *level* of an effect can **not** be set (but can be assigned/controlled, or at least measured), we can treat the effect as a random variable arising from its own distribution.

Random vs. Fixed Effects

$$y_{ij} = \mu + \underbrace{\tau_i}_{\text{fixed}} + \underbrace{\beta_j}_{\text{random}} + \epsilon_{ij}, \quad \beta_j \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_\beta^2), \quad \epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$$

- We're still interested in the treatment effects, τ_i, but not so much β_j directly.
  - Can't set group level, so specific effects aren't useful for future predictions.
  - Knowing σ_β^2 would help us account for extra variation induced by random effect.
  - Want to make inference involving following parameters: τ_i, σ^2, σ_β^2.

Random vs. Fixed Effects

$$\begin{aligned} y_{ij} &= \mu + \tau_i + \beta_j + \epsilon_{ij}, \;\; \epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2) \\ \implies \; y_{ij} | \beta_j &\sim \mathrm{N}(\mu + \tau_i + \beta_j, \sigma^2) \end{aligned}$$

- What is the model for y_{ij} unconditional on β_j?
- Some properties of RVs to recall:

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X|Y]] \\ \mathrm{Var}(X) &= \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathrm{E}[X|Y]) \end{aligned}$$

$$\begin{aligned} \mathrm{E}[y_{ij}] &= \mathrm{E}[\mathrm{E}[y_{ij}|\beta_j]] \\ &= \mathrm{E}[\mu + \tau_i + \beta_j] \\ &= \mu + \tau_i + \underbrace{\mathrm{E}[\beta_j]}_{0} \end{aligned}$$

Random vs. Fixed Effects

$$y_{ij}|\beta_j \sim \mathrm{N}(\mu + \tau_i + \beta_j, \sigma^2)$$

- Some properties of RVs to recall:

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X|Y]] \\ \mathrm{Var}(X) &= \mathrm{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathrm{E}[X|Y]) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(y_{ij}) &= \mathrm{E}[\mathrm{Var}(y_{ij}|\beta_j)] + \mathrm{Var}(\mathrm{E}[y_{ij}|\beta_j]) \\ &= \mathrm{E}[\sigma^2] + \mathrm{Var}(\mu + \tau_i + \beta_j) \\ &= \sigma^2 + \sigma_\beta^2 \end{aligned}$$

(1) Why does $\mathrm{Var}(\mu + \tau_i + \beta_j) = \sigma_\beta^2$?

Random vs. Fixed Effects

- Investigating the covariance between two arbitrary observations.
- Assume $i \neq i'$ and/or $j \neq j'$, because we already worked out variance.

$$
\begin{aligned}
\mathrm{Cov}(y_{ij}, y_{i'j'}) &= \mathrm{E}[y_{ij}y_{i'j'}] - \mathrm{E}[y_{ij}]\mathrm{E}[y_{i'j'}] \\
&= \mathrm{E}\left[\mathrm{E}[y_{ij}y_{i'j'}|\beta_j, \beta_{j'}]\right] - \mathrm{E}[y_{ij}]\mathrm{E}[y_{i'j'}] \\
&= \mathrm{E}[(\mu + \tau_i + \beta_j)(\mu + \tau_{i'} + \beta_{j'})] - (\mu + \tau_i)(\mu + \tau_{i'}) \\
&= \mathrm{E}[(\mu + \tau_i)(\mu + \tau_{i'})] + \mathrm{E}[(\mu + \tau_i)\beta_{j'}] + \mathrm{E}[(\mu + \tau_{i'})\beta_j] + \mathrm{E}[\beta_j\beta_{j'}] - (\mu + \tau_i)(\mu + \tau_{i'}) \\
&= \mathrm{E}[\beta_j\beta_{j'}]
\end{aligned}
$$

(2) Explain how we get from 5 terms to 1 in the final step.

Random vs. Fixed Effects

$$\mathrm{E}[\beta_j \beta_{j'}] = \begin{cases} \sigma_\beta^2, & j = j' \\ 0, & j \neq j' \end{cases}$$

$$\implies \mathrm{Cov}(y_{ij}, y_{i'j'}) = \begin{cases} \sigma^2 + \sigma_\beta^2, & i = i', j = j' \\ \sigma_\beta^2, & i \neq i', j = j' \\ 0, & i \neq i', j \neq j' \end{cases}$$

$$\mathrm{Cov}(\mathbf{y}) = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2 + \sigma_\beta^2 & \sigma_\beta^2 & \dots & \sigma_\beta^2 \\ \sigma_\beta^2 & \sigma^2 + \sigma_\beta^2 & \dots & \sigma_\beta^2 \\ \vdots & & \ddots & \vdots \\ \sigma_\beta^2 & \sigma_\beta^2 & \dots & \sigma^2 + \sigma_\beta^2 \end{pmatrix} = \sigma^2 \mathbf{I}_a + \sigma_\beta^2 \mathbf{J}_a$$

(3) What happens to the **correlation** between y_{ij} and $y_{i'j}$ as $\sigma_\beta^2 \to \infty$? $\to 0$?

Random vs. Fixed Effects

- It takes a bit more math, but it's straightforward to show that the unconditional RVs, y_{ij}, are jointly multivariate normal.
- We have therefore completely characterized their distribution.
- The likelihood function for the parameters is a big multivariate normal density, and can be maximized to obtain (REstricted) Maximum Likelihood estimates.

Implementation

Vascular Example[1]

- Artificial veins produced by manufacturer through extrusion process.
- Rate of defects ("flicks") thought to depend on extrusion pressure, and also possibly the batch of material used.
- Four pressure levels randomly assigned within blocks defined by material batch.

```
vascular <- read.csv("../3_Randomized_Block_Latin_Squares/vascular.csv")
head(vascular)
##   batch pressure flicks
## 1     1     8500   90.3
## 2     1     8700   92.5
## 3     1     8900   85.5
## 4     1     9100   82.5
## 5     2     8500   89.2
## 6     2     8700   89.5
```

[1] DAE Example 4.1 p.144

Vascular Example: Fixed Batch Effects

```
vascular_aov <- aov(flicks ~ as.factor(pressure) + as.factor(batch), data = vascular)
summary(vascular_aov)
##                     Df Sum Sq Mean Sq F value  Pr(>F)
## as.factor(pressure)  3  178.2   59.39   8.107 0.00192 **
## as.factor(batch)     5  192.2   38.45   5.249 0.00553 **
## Residuals           15  109.9    7.33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vascular Example

$$y_{ij} = \mu + \underbrace{\tau_i}_{\text{fixed}} + \underbrace{\beta_j}_{\text{random}} + \epsilon_{ij}, \quad \beta_j \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2_\beta), \quad \epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$$

- Interested in studying τ_i.
- Include $\beta_j \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2_\beta)$ to:
  - reduce mean squared error and therefore increase our power to detected small treatment effects and
  - account for positive dependence among observations that share a nuisance factor level.
- The R formula syntax for this model is `y ~ treatment + (1 | blocks)` in the `lme4` package.
- β_j is called a "random intercept".

Vascular Example: `lme4::lmer()`

```
library(lme4)
## Loading required package: Matrix
vascular_RE <- lmer(flicks ~ as.factor(pressure) + (1 | batch), data = vascular)
anova(vascular_RE)
## Analysis of Variance Table
##                     npar Sum Sq Mean Sq F value
## as.factor(pressure)    3 178.17   59.39  8.1071
confint(vascular_RE)
## Computing profile confidence intervals ...
##                             2.5 %     97.5 %
## .sig01                   1.136322  5.5401925
## .sigma                   1.840242  3.5643700
## (Intercept)             89.726101 95.9072331
## as.factor(pressure)8700 -4.085262  1.8185951
## as.factor(pressure)8900 -6.851928 -0.9480715
## as.factor(pressure)9100 -10.001928 -4.0980715
```

- Degrees of freedom values get trickier with mixed models, so `lme4` opts to simply withhold p-values.

Vascular Example: `lmerTest::lmer()`

- The `lmerTest` package implements a method for computing degrees of freedom (Satterthwaite's method).
- Another option (not shown here) is to use bootstrapping.[2]

```
library(lmerTest)
##
## Attaching package: 'lmerTest'
## The following object is masked from 'package:lme4':
##
##     lmer
## The following object is masked from 'package:stats':
##
##     step
vascular_RE2 <- lmer(flicks ~ as.factor(pressure) + (1 | batch), data = vascular)
```

```
anova(vascular_RE2)
## Type III Analysis of Variance Table with Satterthwaite's method
##                     Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## as.factor(pressure) 178.17   59.39     3    15  8.1071 0.001916 **
```

[2] see `?bootMer()` and the boot package to learn more

Vascular Example: `lmerTest::lmer()`

```
summary(vascular_RE2)
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  batch    (Intercept) 7.781    2.789
##  Residual             7.326    2.707
## Number of obs: 24, groups:  batch, 6
##
## Fixed effects:
##                         Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)               92.817      1.587 11.136  58.494 3.24e-15 ***
## as.factor(pressure)8700   -1.133      1.563 15.000  -0.725 0.479457
## as.factor(pressure)8900   -3.900      1.563 15.000  -2.496 0.024713 *
## as.factor(pressure)9100   -7.050      1.563 15.000  -4.512 0.000414 ***
```

(4) Can you find the MSE estimate for σ^2? Estimate of σ^2_β?

Vascular Example: `lme4::lmer()`

```
ranova(vascular_RE2)
## ANOVA-like table for random-effects: Single term deletions
##
## Model:
## flicks ~ as.factor(pressure) + (1 | batch)
##              npar  logLik    AIC    LRT Df Pr(>Chisq)
## <none>          6 -56.021 124.04
## (1 | batch)     5 -59.114 128.23 6.1853  1    0.01288 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```