Intro to mixed effects models

- This week:
  - Monday: intro to mixed effects models
  - Wednesday: more mixed effects models, Course evaluations
  - Friday: wrap up mixed effects models
- HW 7 due Friday
- Exam 2 resubmission due Friday, Dec 9 at 12pm
- No final exam!
- Class dinner during finals week?

Motivating example: performance anxiety

We have data from a 2010 study on performance anxiety in 37 undergraduate music majors. For each musician, data was collected on anxiety levels before different performances (between 2 and 15 performances were measured for each musician), with variables including:

- `id`: a unique identifier for the musician
- `na`: negative affect score (a measure of anxiety)
- `perform_type`: whether the musican was performing in a large ensemble, small ensemble, or solo

> How can we model the relationship between performance type and anxiety?

A linear model for anxiety

$$Anxiety_i = \beta_0 + \beta_1\, SmallEnsemble_i + \beta_2\, Solo_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$$

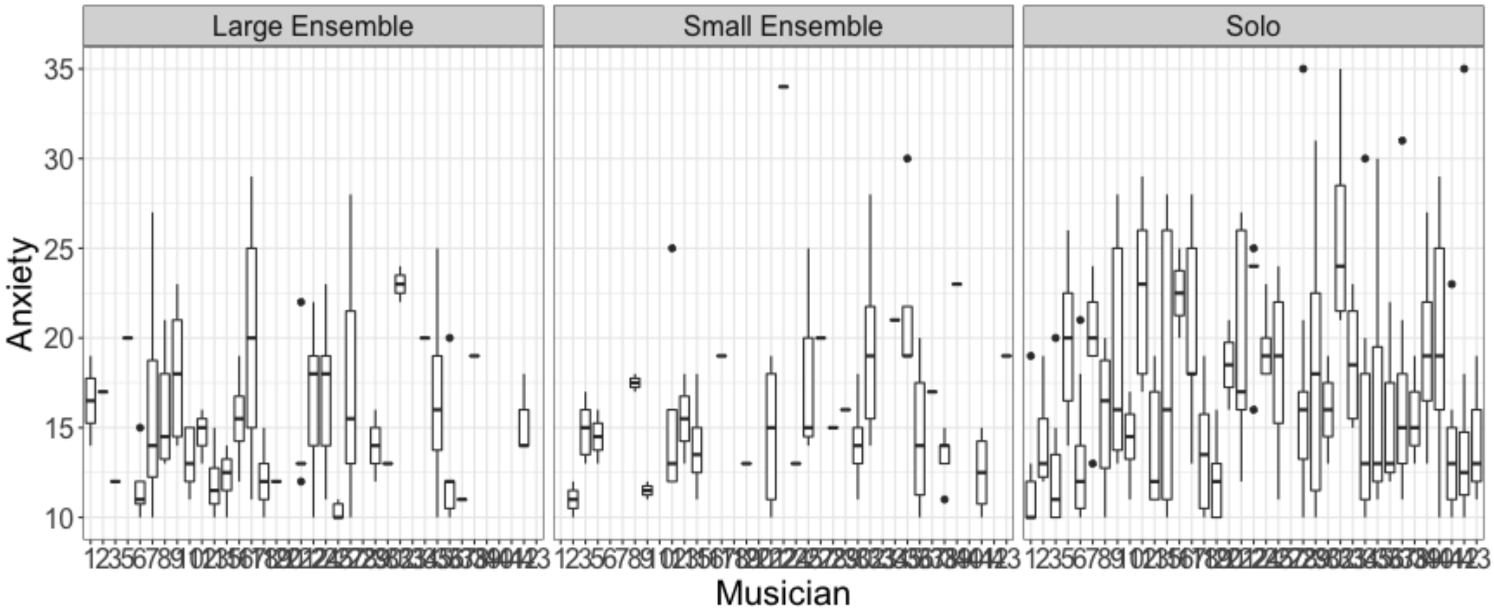
> What assumptions does this linear model make? Are all the assumptions reasonable?

- independence: anxiety scores for different performances are independent
  - not reasonable, b/c we have multiple performances for each musician!

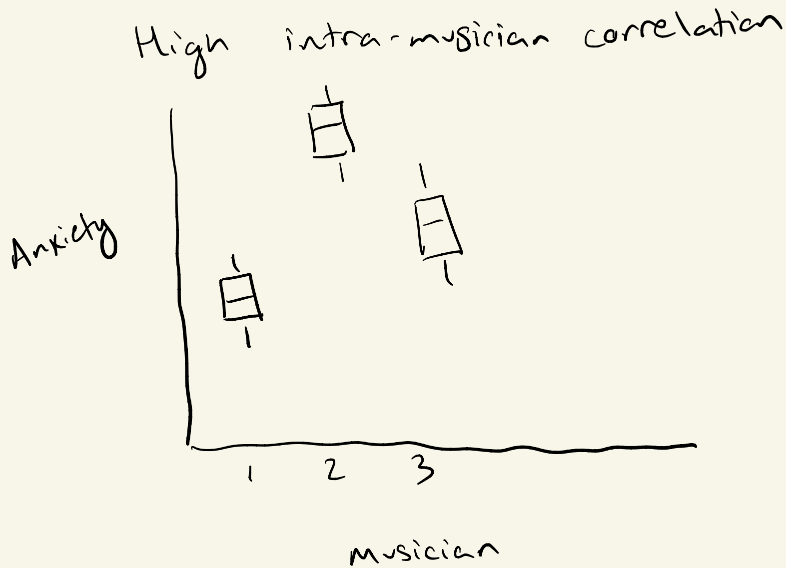
Repeated measures data: multiple observations from an individual (e.g. a patient)

- normality
- constant variance

Exploratory data analysis



> Does it look like anxiety is correlated within musicians?



between-group variance = variability from musician to musician

within-group variance = variability between performances for the same musician

high intra-musician correlation occurs when between-group variance is high, relative to within-group variance

Changing the model

$$Anxiety_i = \beta_0 + \beta_1 \, SmallEnsemble_i + \beta_2 \, Solo_i + \varepsilon_i$$

$$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$$

Option 2: have a different distribution for ε that allows dependence

How can we change the model to account for correlation within musicians?

Option 1: $Anxiety_i = \beta_0 + \beta_1 Small_i + \beta_2 Solo_i + \beta_3 Musician2_i + \beta_4 Musician3_i + \cdots + \beta_{38} Musician37_i + \varepsilon_i$

Problems:

1) Lots of βs to estimate!
2) Not much data to estimate some βs
3) Doesn't generalize to population
4) we don't care about the coefficients for each musician

A mixed effects model

fixed effects

$$Anxiety_{ij} = \beta_0 + u_i + \beta_1 \; SmallEnsemble_{ij} + \beta_2 \; Solo_{ij} + \varepsilon_{ij}$$

noise

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2) \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

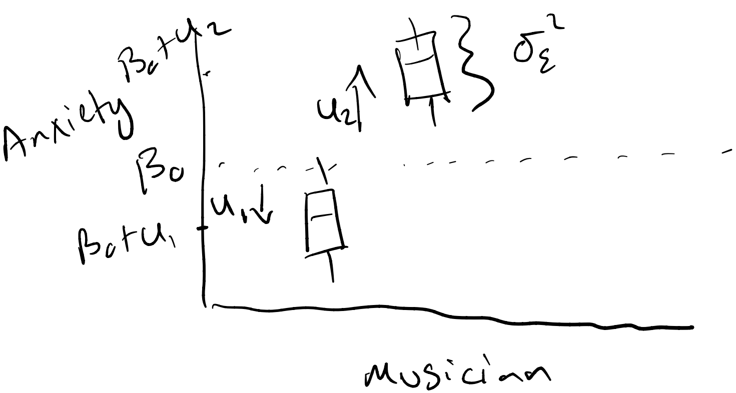
$Anxiety_{ij}$ = anxiety of musician i before performance j

u_i = random effect for musician i
(random variable, not parameter)

$\beta_0 + u_i$ = intercept for musician i

σ_u^2 = variability between musicians

σ_ε^2 = variability between performances within a musician



$$\text{Anxiety}_{ij} = \beta_0 + u_i + \beta_1 \text{Small}_{ij} + \beta_2 \text{Solo}_{ij} + \varepsilon_{ij} \qquad u_i \overset{iid}{\sim} N(0, \sigma_u^2)$$

Matrix form: $Y = X\beta + Zu + \varepsilon$ ← noise

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \end{bmatrix} \sim N(0, \sigma_\varepsilon^2 I)$$

(length = # obs. e.g. 497)

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \end{bmatrix}$$

X = design matrix for fixed effects

Z = design matrix for random effects (which group / musician does each row come from)

E.g.,

$$X = \begin{bmatrix} 1 & \text{Small}_{11} & \text{Solo}_{11} \\ 1 & \text{Small}_{12} & \text{Solo}_{12} \\ \vdots & \vdots & \vdots \\ 1 & \vdots & \vdots \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & \cdots & 1 \end{bmatrix} \begin{matrix} \left.\right\} \text{group 1 (musician 1)} \\ \\ \left.\right\} \text{group 2 (musician 2)} \\ \\ \end{matrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \end{bmatrix} \sim N(0, \sigma_u^2 I)$$

length = # groups in data (e.g. 37 musicians)

$$\begin{aligned} \text{Var}(Y) &= \text{Var}(Zu) + \text{Var}(\varepsilon) \\ &= Z(\sigma_u^2 I)Z^T + \sigma_\varepsilon^2 I \end{aligned}$$

$$
\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{Var}(Zu) + \mathrm{Var}(\varepsilon) \\
&= Z(\sigma_u^2 I)Z^T + \sigma_\varepsilon^2 I
\end{aligned}
$$

$$
= \begin{bmatrix}
\left.\begin{matrix}
\sigma_u^2+\sigma_\varepsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\
\sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & & \sigma_u^2 \\
\vdots & & \ddots & \vdots \\
\sigma_u^2 & & & \sigma_u^2+\sigma_\varepsilon^2
\end{matrix}\right\} \text{group 1 (musician 1)} & 0 \\
0 & \text{group 2 (musician 2)} \left\{\begin{matrix}
\sigma_u^2+\sigma_\varepsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\
\sigma_u^2 & \sigma_u^2+\sigma_\varepsilon^2 & & \\
\vdots & & \ddots &
\end{matrix}\right.
\end{bmatrix}
$$

Correlation within a group (aka "intra-class correlation") ICC :

$$
\frac{\overset{\text{between-group variance}}{\sigma_u^2}}{\underbrace{\sigma_u^2+\sigma_\varepsilon^2}_{\text{total variance}}}
$$

within-group variance

Fitting the model in R

```
library(lme4)
m1 <- lmer(na ~ perform_type + (1|id),
           data = music)
summary(m1)
```

```
...
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept)  5.56    2.358
##  Residual             21.75    4.664
## Number of obs: 497, groups:  id, 37
## Fixed effects:
##                              Estimate Std. Error t value
## (Intercept)                   14.9654     0.5920  25.278
## perform_typeSmall Ensemble     0.7709     0.7210   1.069
## perform_typeSolo               2.0142     0.5521   3.648
...
```

Assumptions

$$Anxiety_{ij} = \beta_0 + u_i + \beta_1 \ SmallEnsemble_{ij} + \beta_2 \ Solo_{ij} + \varepsilon_{ij}$$

$$u_i \overset{iid}{\sim} N(0, \sigma_u^2) \qquad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma_\varepsilon^2)$$

> What assumptions does this mixed effects model make?