Unlocking the Multilevel Equation

Multicollinearity, Effect Size, Design Effects, and Power Analysis

Mark Lai

University of Southern California

2024-03-22

Overview

- The mixed model equation
- Multicollinearity and variance partitioning (with Venn diagrams)
- Standardized effect size
- Design effect/variance inflation
- Power analysis

Slides available at https://quantscience.rbind.io/presentation

Mixed Model Equation

Multilevel Data

cid	y	x1	x2
<int>	<int>	<dbl>	<int>
1	10	0	2
1	5	0	0
1	5	0	1
1	7	0	2
2	5	1	2
2	5	1	2
2	5	1	2
2	8	1	1
3	8	1	0
3	5	1	0

1-10 of 16 rows — Previous **1** 2 Next

Mixed Model Equations

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\begin{bmatrix} 10 \\ 5 \\ 5 \\ 7 \\ 5 \\ 5 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} + \begin{bmatrix} 1 & & 2 & \\ 1 & & 0 & \\ 1 & & 1 & \\ 1 & & 2 & \\ & 1 & & 1 \\ & 1 & & 2 \\ & 1 & & 2 \\ & 1 & & 2 \end{bmatrix} \begin{bmatrix} u_{01} \\ u_{02} \\ u_{11} \\ u_{12} \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \end{bmatrix}$$

For cluster j, $\mathbf{y}_j = \mathbf{X}_j\gamma + \mathbf{Z}_j\mathbf{u}_j + \mathbf{e}_j$

Mixed Model Equations (cont'd)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

- $\mathbf{Z} = [\mathbf{Z}_0 \cdots \mathbf{Z}_{q-1}]$, and each component is block-diagonal
- $\mathbf{e}$ is a vector of errors;[1] $V(\mathbf{e}) = \sigma^2\mathbf{I}$
- $\mathbf{u}$: vector of length qJ of random effects; $V(\mathbf{u}_j) = \mathbf{T} = \sigma^2\mathbf{D} \quad \forall j$

$$\mathbf{D} = \begin{bmatrix} \tau_0^2/\sigma^2 & & & \\ \tau_{01}/\sigma^2 & \tau_1^2/\sigma^2 & & \\ \vdots & \cdots & \ddots & \\ \tau_{0q-1}/\sigma^2 & \tau_{1q-1}/\sigma^2 & \cdots & \tau_{q-1}^2/\sigma^2 \end{bmatrix}$$

- $\mathbf{u}$ and $\mathbf{e}$ are mutually independent, and both are indepedent of $\mathbf{X}$ and $\mathbf{Z}$

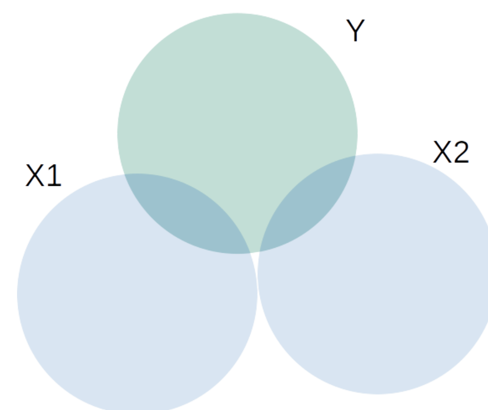
Multicollinearity

Variance Partitioning

In regression, we look at

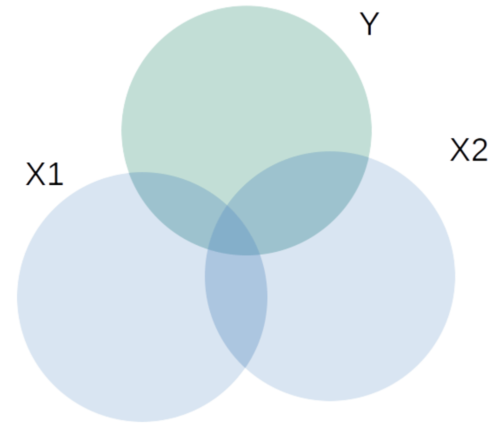
- $\mathbf{X}^{c\top}\mathbf{X}^c/N$, covariance among columns of $\mathbf{X}^c$ (assume mean centered predictors)

If X_1 and X_2 are uncorrelated, $\mathbf{x}_1^{c\top}\mathbf{x}_2^c = 0$, and variance accounted for by the two predictors are separate



If X_1 and X_2 are correlated, $\mathbf{x}_1^{c\top}\mathbf{x}_2^c \neq 0$, then adding $X_2 \ldots$

- changes the coefficient of X_1, and
- increases the standard error of the coefficient



Rearranging the MLM Equation

$$
\begin{bmatrix} 10 \\ 5 \\ 5 \\ 7 \\ 5 \\ 5 \\ 5 \\ 8 \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 2 & 1 & & 2 & \\
1 & 0 & 0 & 1 & & 0 & \\
1 & 0 & 1 & 1 & & 1 & \\
1 & 0 & 2 & 1 & & 2 & \\
1 & 1 & 2 & & 1 & & 1 \\
1 & 1 & 2 & & 1 & & 2 \\
1 & 1 & 2 & & 1 & & 2 \\
1 & 1 & 1 & & 1 & & 2
\end{bmatrix}
\begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ u_{01} \\ u_{02} \\ u_{11} \\ u_{12} \end{bmatrix}
+
\begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{24} \end{bmatrix}
$$

In MLM, we also look at

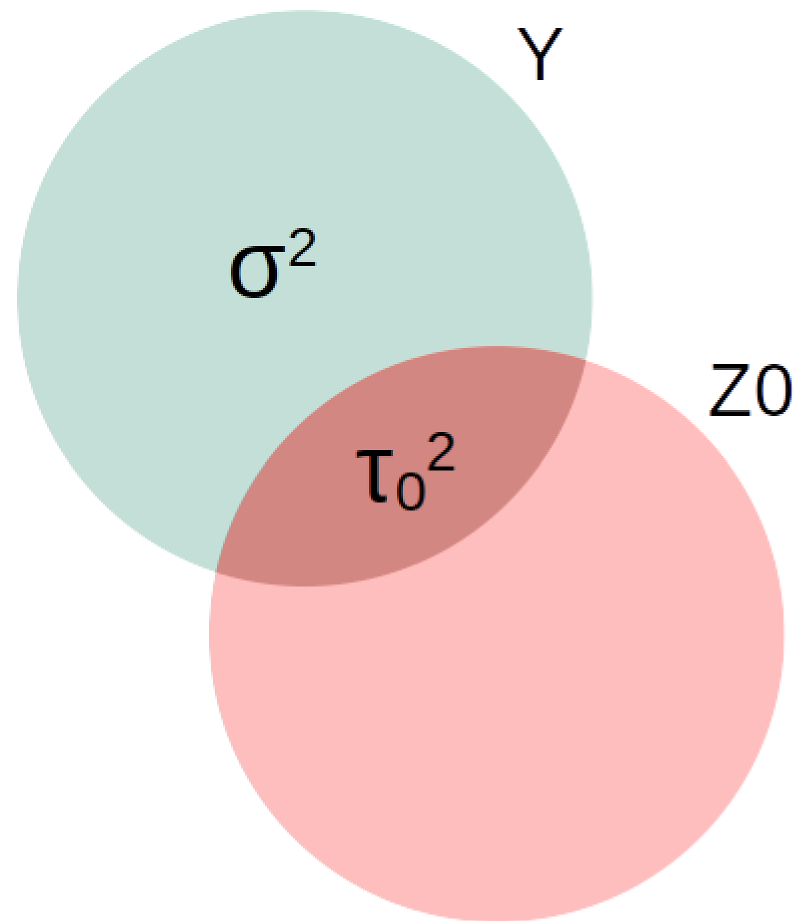
- $\mathbf{Z}^\top \mathbf{Z}$, cross-products among columns of $\mathbf{Z}$, and
- $\mathbf{X}^{c\top} \mathbf{Z}$, cross-products between columns of $\mathbf{X}^c$ and those of $\mathbf{Z}$

Example 1: Random Intercepts Only

- $\mathbf{Z}_0 = \mathrm{diag}(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_J}) \Rightarrow \mathbf{x}_0^\top \mathbf{Z}_0 \neq \mathbf{0}$

```
1  Z0
```

```
     [,1] [,2]
[1,]    1    0
[2,]    1    0
[3,]    1    0
[4,]    0    1
[5,]    0    1
[6,]    0    1
```



- Residual maker matrix: $\mathbf{M}_{Z_0} = \mathbf{I} - \mathbf{Z}_0(\mathbf{Z}_0^\top \mathbf{Z}_0)^{-1}\mathbf{Z}_0^\top$
  - $\mathbf{M}_{Z_0}\mathbf{Z}_0 = \mathbf{0}$

```r
(M_Z0 <- diag(nrow(Z0)) - Z0 %*% solve(crossprod(Z0), t(Z0))) |>
    round(digits = 2)
```

```
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,]  0.67 -0.33 -0.33  0.00  0.00  0.00
[2,] -0.33  0.67 -0.33  0.00  0.00  0.00
[3,] -0.33 -0.33  0.67  0.00  0.00  0.00
[4,]  0.00  0.00  0.00  0.67 -0.33 -0.33
[5,]  0.00  0.00  0.00 -0.33  0.67 -0.33
[6,]  0.00  0.00  0.00 -0.33 -0.33  0.67
```

Example 2a: Level-2 Predictor

```
[1] -1 -1 -1  1  1  1
```

- $\mathbf{x}_1 = \left[w_1 \mathbf{1}_{n_1}^\top, \ldots, w_J \mathbf{1}_{n_J}^\top\right]^\top \Rightarrow \mathbf{x}_1^\top \mathbf{Z}_0 \neq \mathbf{0}$

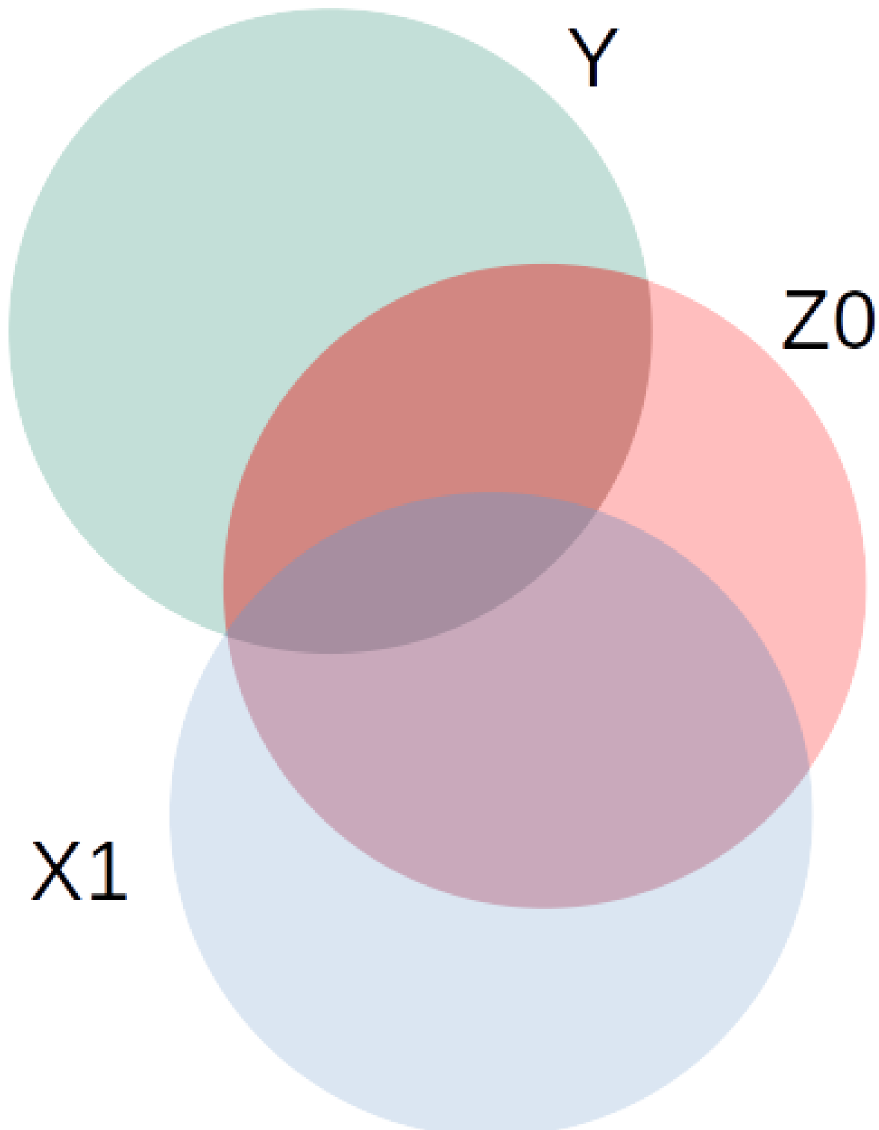
```r
crossprod(X1, Z0)
```

```
     [,1] [,2]
[1,]   -3    3
```

- $\mathbf{x}_1^\top \mathbf{M}_{Z_0} = \mathbf{0}$

```r
crossprod(X1, M_Z0) |>
    round(digits = 2)
```

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    0    0    0    0    0    0
```



Level-2 predictor can only account for level-2 variance

Example 2b: Pure Level-1 Predictor

```
[1] -2  0  2 -3  2  1
```

- Cluster means of X_1 = 0 $\Rightarrow \mathbf{x}_1^\top \mathbf{Z}_0 = \mathbf{0}$

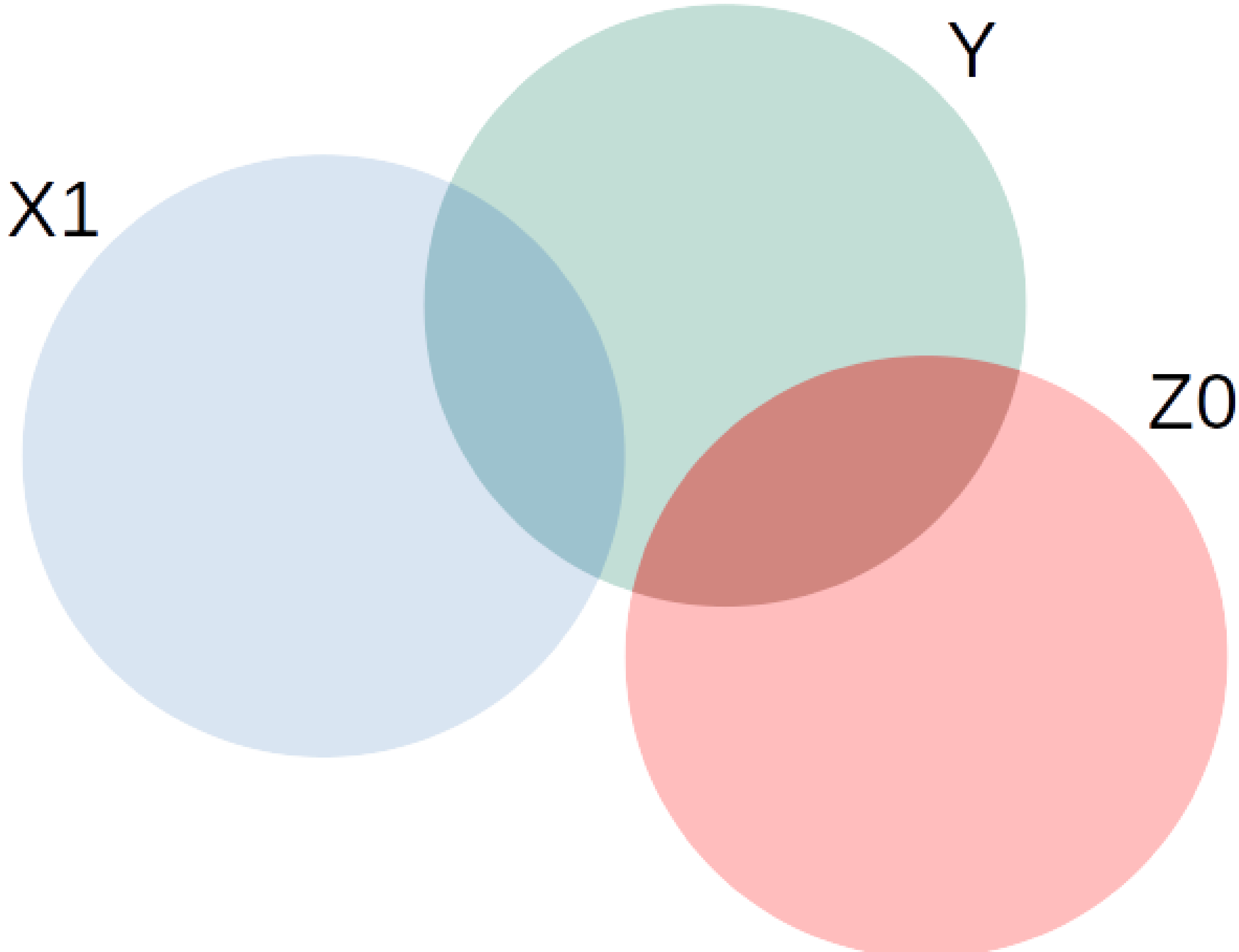
```r
crossprod(X1, Z0)
```

```
     [,1] [,2]
[1,]    0    0
```

- $\mathbf{x}_1^\top \mathbf{M}_{Z_0} \neq \mathbf{0}$

```r
crossprod(X1, M_Z0)
```

```
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]   -2    0    2   -3    2    1
```



Pure level-1 predictor can only account for level-1 variance

Example 2c: General Level-1 Predictor

```
[1] -1  1  2 -3  0  1
```

- Cluster means of $\mathbf{x}_1 \neq 0 \Rightarrow \mathbf{x}_1^\top \mathbf{Z}_0 \neq \mathbf{0}$

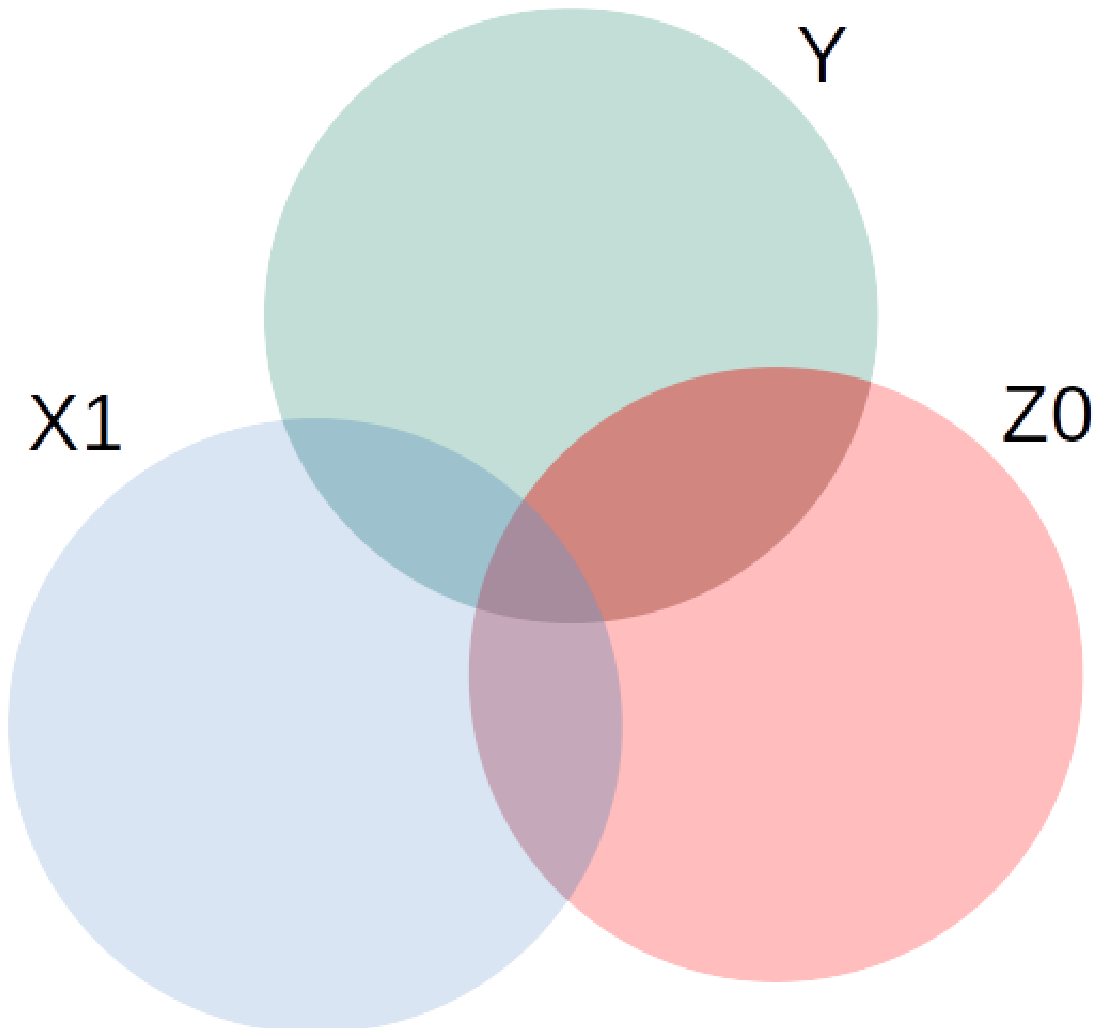
```r
crossprod(X1, Z0)
```

```
     [,1] [,2]
[1,]    2   -2
```

- $\mathbf{x}_1^\top \mathbf{M}_{Z_0} \neq \mathbf{0}$

```r
crossprod(X1, M_Z0) |>
    round(digits = 2)
```

```
      [,1] [,2] [,3]  [,4] [,5] [,6]
[1,] -1.67 0.33 1.33 -2.33 0.67 1.67
```



> Level-1 predictor can account for both level-2 and level-1 variance

Example 3: Random slopes

- Without cluster-mean centering, $\mathbf{Z}_1 = \text{diag}(\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1J})$

```
     [,1] [,2]
[1,]   -1    0
[2,]    1    0
[3,]    2    0
[4,]    0   -3
[5,]    0    0
[6,]    0    1
```

- With nonzero cluster means, $\mathbf{Z}_1^\top \mathbf{Z}_0 \neq \mathbf{0}$ and $\mathbf{Z}_1^\top \mathbf{M}_{Z_0} \neq \mathbf{0}$

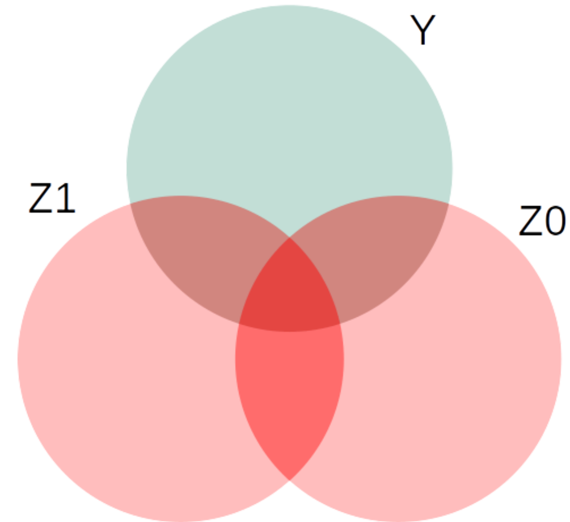
```
crossprod(Z1, Z0)
```

```
     [,1] [,2]
[1,]    2    0
[2,]    0   -2
```

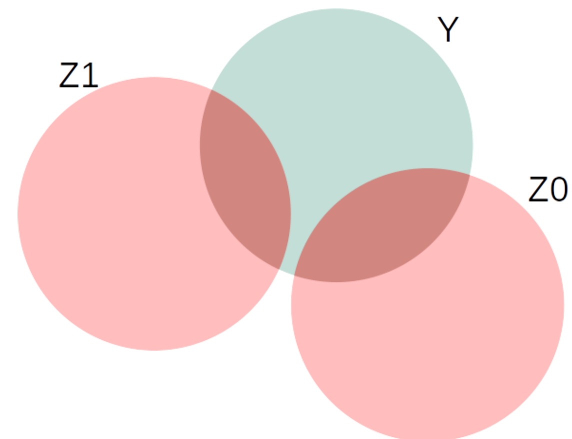
```
t(M_Z0 %*% Z1) |> round(di
```

```
      [,1] [,2] [,3]  [,4] [,5] [,6]
[1,] -1.67 0.33 1.33  0.00 0.00 0.00
[2,]  0.00 0.00 0.00 -2.33 0.67 1.67
```

- For general level-1 predictor
  - Random slope variance (τ_1^2), if omitted, will be redistributed to the fixed effect, random intercept variance (τ_0^2), and σ^2



- For purely level-1 predictor, $\mathbf{Z}_1^\top \mathbf{Z}_0 = \mathbf{0}$
  - Random slope variance (τ_1^2), if omitted, will be redistributed to only fixed effect and σ^2



Revisiting Lai & Kwok (2015)

Using simulations, we found that, when clustering is ignored, SEs of the following are underestimated

- Lv-2 predictor
- General Lv-1 predictor (i.e., with unequal cluster means)
- Lv-1 predictor with random slopes

> Because they all correlate with $\mathbf{Z}$!

Additional Questions

- Can we use this framework to quantify model misspecification?
  - How does omitting components of X affect estimates of γ and τ?
  - How does omitting components of Z affect estimates of γ and τ?
- How do we generalize this beyond two-level models?
  - E.g., Crossed random effects (e.g., Lai, 2019)[1]

Effect Size

Total Variance of $\mathbf{y}$

Conditional on $\mathbf{X}$

$$V(\mathbf{Zu} + \mathbf{e}) = \sigma^2 \mathbf{V} = \sigma^2 [\mathbf{Z}(\mathbf{D} \otimes \mathbf{I}_J)\mathbf{Z}^\top + \mathbf{I}],$$

Unconditional ($\mathbf{X}^c$ = grand-mean centered $\mathbf{X}$)

$$\underbrace{\mathbf{X}^c \boldsymbol{\gamma}\boldsymbol{\gamma}^\top \mathbf{X}^{c\top}}_{\text{fixed}} + \underbrace{\sigma^2 \mathbf{Z}(\mathbf{D} \otimes \mathbf{I}_J)\mathbf{Z}^\top}_{\text{random lv-2}} + \underbrace{\sigma^2 \mathbf{I}}_{\text{random lv-1}}.$$

Average Variance of Each Observation

- Random slopes imply nonconstant variance across observations[1]
  - $V(x_{ij}u_j)$ depends on x_{ij}

```r
sigma2 <- 0.5
D_mat <- diag(c(1, 0.5))
Z <- cbind(Z0, Z1)
(Vy <- sigma2 * (Z %*% (D_
    diag(num_clus)) %*% t
    diag(num_obs * num_clu
```

```
     [,1] [,2] [,3] [,4]  [,5] [,6]
[1,] 1.25 0.25 0.00
[2,] 0.25 1.25 1.00
[3,] 0.00 1.00 2.00
[4,]                3.25 0.50 -0.25
[5,]                0.50 1.00  0.50
[6,]               -0.25 0.50  1.25
```

Average Variance of Each Observation (cont'd)

$$
\begin{aligned}
\mathrm{Tr}[V(\mathbf{y})]/N &= \boldsymbol{\gamma}^\top \mathbf{X}^{c\top} \mathbf{X}^c \boldsymbol{\gamma}/N \\
&\quad + \sigma^2\, \mathrm{Tr}[\mathbf{Z}(\mathbf{D} \otimes \mathbf{I}_J)\mathbf{Z}^\top]/N + \sigma^2
\end{aligned}
$$

```r
mean(diag(Vy))  # just the random component
```

```
[1] 1.666667
```

```
     [,1] [,2] [,3] [,4]  [,5] [,6]
[1,] 1.25 0.25 0.00
[2,] 0.25 1.25 1.00
[3,] 0.00 1.00 2.00
[4,]                3.25 0.50 -0.25
[5,]                0.50 1.00  0.50
[6,]               -0.25 0.50  1.25
```

Using Summary Statistics

$$\mathrm{Tr}[V(\mathbf{y})]/N = \boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_X \boldsymbol{\gamma} + \sigma^2[\mathrm{Tr}(\mathbf{D}\mathbf{K}_Z) + 1],$$

Let $\mathbf{X}^r$ be the design matrix of variables with random effects[1]

- $\boldsymbol{\Sigma}_X = \mathbf{X}^{c\top}\mathbf{X}^c/N$ (covariance matrix of $\mathbf{X}$)
- $N\mathbf{K}_Z = \mathbf{X}^{r\top}\mathbf{X}^r$ (uncentered crossproduct matrix)
  - $\mathbf{K}_Z = \bar{\mathbf{X}}^{r\top}\bar{\mathbf{X}}^r + \boldsymbol{\Sigma}_X^r$

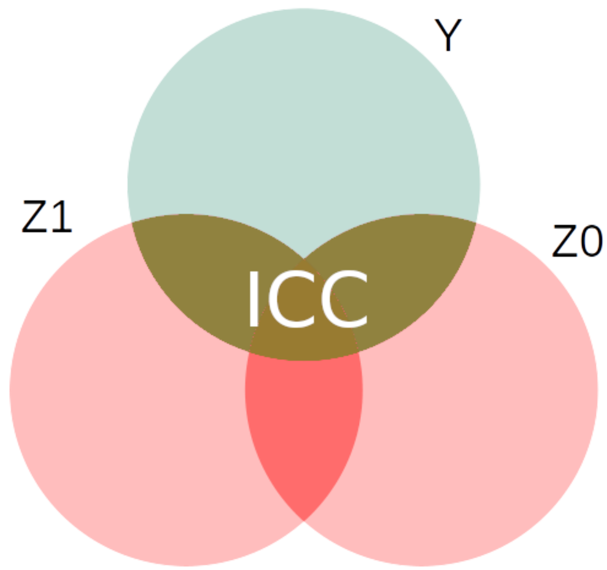
cf. Rights & Sterba (2019); Rights & Sterba (2023)

Intraclass Correlations

Unconditional or Conditional on the fixed effects

- i.e., proportion of variance at level 2 out of all variance unaccounted for by the fixed effects

$$\mathrm{ICC}_D = \frac{\mathrm{Tr}(\mathbf{DK}_Z)}{\mathrm{Tr}(\mathbf{DK}_Z) + 1},$$



Can also be further partitioned by specific components of $\mathbf{Z}$ (e.g., random slopes, see

Multilevel R^2

Proportion of variance by fixed effects (see Johnson, 2014; Rights & Sterba, 2019)[1]

$$R^2_{\text{GLMM}} = \frac{\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_X \boldsymbol{\gamma}}{\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_X \boldsymbol{\gamma} + \sigma^2 [\text{Tr}(\mathbf{D}\mathbf{K}_Z) + 1]}.$$

Code Example

```
Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ ses + meanses + (ses | school)
   Data: Hsb82

REML criterion at convergence: 46561.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1671 -0.7269  0.0163  0.7547  2.9646

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept)  2.695   1.6418
          ses          0.453   0.6731   -0.21
 Residual             36.796   6.0659
Number of obs: 7185, groups:  school, 160

Fixed effects:
            Estimate Std. Error t value
(Intercept)  12.6740     0.1506  84.176
ses           2.1903     0.1218  17.976
meanses       3.7812     0.3826   9.882
```

```r
# Variance by fixed effects
sigmax <- cov(Hsb82[c("ses", "meanses")]) /
    nrow(Hsb82) * (nrow(Hsb82) - 1)
(vf <- crossprod(fixef(m1)[2:3],
                 sigmax %*% fixef(m1)[2:3])
```

```
         [,1]
[1,] 8.190918
```

```r
# Variance by Z
sigmaz <- matrix(c(0, 0, 0, sigmax[1, 1]),
Kz <- sigmaz + tcrossprod(c(1, mean(Hsb82$s
tau_mat <- VarCorr(m1)[[1]]
(vr <- sum(Kz * tau_mat))
```

```
[1] 2.970493
```

Code Example (cont'd)

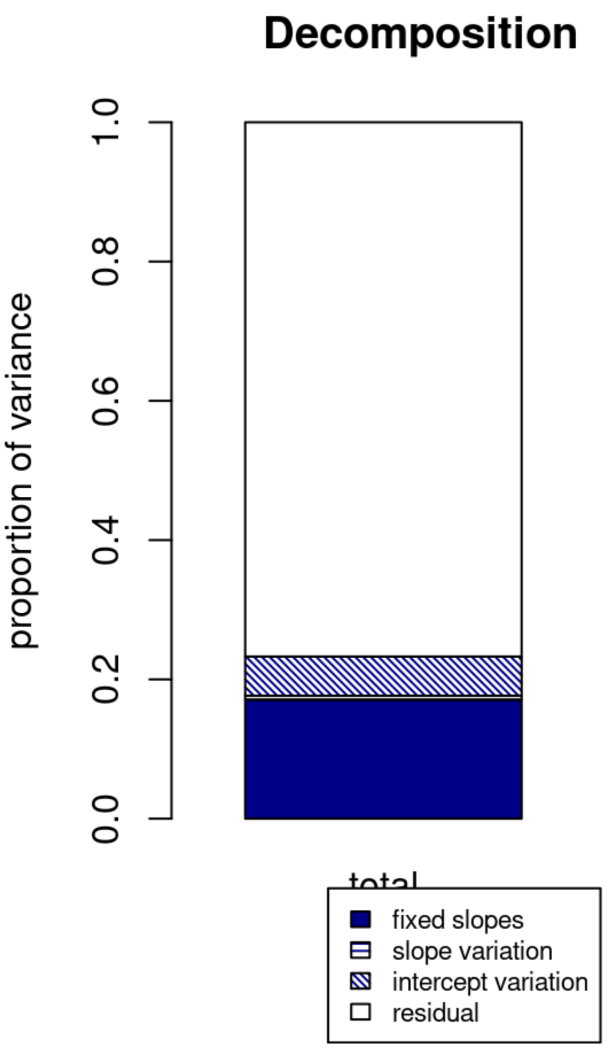
```r
library(r2mlm)
library(MuMIn)
# R^2 using formula, r2mlm::r2mlm(), and MuMIn::r.squar
list(formula = vf / (vf + vr + sigma(m1)^2),
     r2mlm = r2mlm(m1),
     MuMIn = MuMIn::r.squaredGLMM(m1))
```

```
$formula
         [,1]
[1,] 0.170797

$r2mlm
$r2mlm$Decompositions
                      total
fixed           0.170816580
slope variation 0.005737776
mean variation  0.056202223
sigma2          0.767243421

$r2mlm$R2s
          total
f   0.170816580
v   0.005737776
m   0.056202223
fv  0.176554356
fvm 0.232756579


$MuMIn
```



Confidence interval of R^2

- Delta method
- Bootstrapping (with `lme4::bootMer()` or R package *bootmlm*)[1]
- Bayesian estimation

Additional Questions

- What is the bias of sample R^2 estimator?
  - Can we compute adjusted R^2?

Standardized Mean Difference

SMD = Mean difference / SD_p[1]

For cluster-randomized trials

$$Y_{ij} = T_j\gamma + Z_{0j}u_{0j} + e_{ij}, \quad V(u_{0j}) = \tau_0^2$$

T_j = 0 (control) and 1 (treatment)

$$\delta = \frac{\gamma}{\sqrt{\sigma^2[\mathrm{Tr}(\mathbf{DK}_Z) + 1]}} = \frac{\gamma}{\sqrt{\tau_0^2 + \sigma^2}}$$

Sample estimator d: replace γ, τ_0^2, and σ^2 with sample estimates

More general form:

$$\delta = \frac{\text{adjusted/unadjusted treatment effect}}{\sqrt{\text{average conditional variance of } \mathbf{y}}}$$

$$\delta = \frac{E[E(Y|T=1) - E(Y|T=0)|\mathbf{C}]}{\sqrt{\mathrm{Tr}[V(\mathbf{X}\boldsymbol{\gamma}|T) + V(\mathbf{Zu}|T)]/N + \sigma^2}}$$

where $\mathbf{C}$ is the subset of covariates for adjusted differences

Linking R^2 to δ

$$f^2_{\text{GLMM}} = \frac{R^2_{\text{GLMM}}}{1 - R^2_{\text{GLMM}}}$$

- For two-group designs with one predictor, $2f = d$

Additional Questions

- What ranges of effect sizes do published MLM studies have?

Standardized Coefficients

$$\gamma^s = \gamma \frac{SD_x}{\sqrt{\mathbf{Tr}[V(\mathbf{y})]/N}}$$

```
Linear mixed model fit by REML ['lmerMod']
Formula: mAch ~ ses + meanses + (ses | school)
   Data: Hsb82

REML criterion at convergence: 46561.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1671 -0.7269  0.0163  0.7547  2.9646

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept)  2.695   1.6418
          ses          0.453   0.6731   -0.21
 Residual             36.796   6.0659
Number of obs: 7185, groups:  school, 160
```

For cluster means and cluster-mean centered variables, it is still natural to use the total SD

- SD_{SES} = 0.78
- $\boldsymbol{\Sigma}_X$ = $\begin{bmatrix} 0.61 & 0.17 \\ 0.17 & 0.17 \end{bmatrix}$
- $\mathbf{K}_Z = \begin{bmatrix} 1 & 0 \\ 0 & 0.61 \end{bmatrix}$, $\sigma^2\mathbf{D}$ = $\begin{bmatrix} 2.7 & -0.23 \\ -0.23 & 0.45 \end{bmatrix}$
- $\mathrm{Tr}[V(\mathbf{y})]/N$ = $\boldsymbol{\gamma}^\top \boldsymbol{\Sigma}_X \boldsymbol{\gamma}$ + $\sigma^2[\mathrm{Tr}(\mathbf{D}\mathbf{K}_Z) + 1]$ = 8.19 + 2.97 + 36.8 = 47.96
- γ^s_{ses} = 2.19 / $\sqrt{47.96}$ = 0.25; $\gamma^s_{\text{meanses}}$ = 3.78 / $\sqrt{47.96}$ = 0.43

Design Effect

Design Effect/Variance Inflation

Design effect: Expected impact of design on sampling variance of an estimator[1]

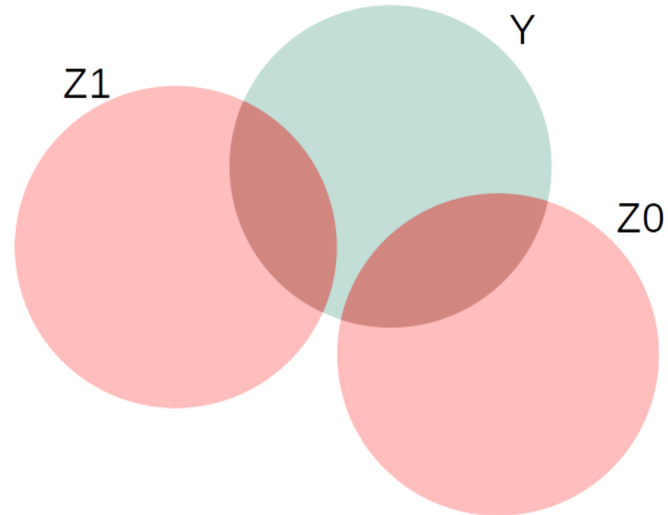
For the sample mean ($\hat{\mu}$):

- Simple random sample: $V_{\text{SRS}}(\hat{\mu}_{\text{SRS}}) = \tilde{\sigma}^2/N$
  - *Assuming constant total variance*: $\tilde{\sigma}^2 = \tau_0^2 + \sigma^2$
- With clustering:[2] $V_{\text{MLM}}(\hat{\gamma}_0) = \tau_0^2/J + \sigma^2/N$

$$\text{Deff}(\hat{\mu}) = \frac{V_{\text{SRS}}(\hat{\gamma}_0)}{V_{\text{MLM}}(\hat{\mu}_{\text{SRS}})} = 1 + (n-1)\text{ICC}$$

ⓘ $\mathrm{Deff}(\hat{\mu})$ Does Not Inform About Random Slopes

- Even when $\mathrm{Deff}(\hat{\mu})$ is close to 1, random slope variance τ_1^2 can still be large
  - Random slopes can be independent from random intercepts




```r
# Without including random slopes, it seems ICC = 0, Deff = 0
m0 <- lmer(y ~ (1 | clus_id), data = dat)
VarCorr(m0)
```

```
Groups   Name        Std.Dev.
clus_id  (Intercept) 0.0000
Residual             1.0798
```

```r
# But the random slope variance can be quite large
m1 <- lmer(y ~ x + (x | clus_id), data = dat)
VarCorr(m1)
```

```
Groups   Name        Std.Dev. Corr
clus_id  (Intercept) 0.20239
         x           0.54036  0.704
Residual             0.93972
```

Deff for $\hat{\gamma}$

$$\hat{\boldsymbol{\gamma}} = (\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{y}$$

$$V_{\mathrm{MLM}}(\hat{\boldsymbol{\gamma}}^{\mathrm{GLS}}) = \sigma^2(\mathbf{X}^\top \mathbf{V}^{-1}\mathbf{X})^{-1}$$
$$V_{\mathrm{SRS}}(\hat{\boldsymbol{\gamma}}^{\mathrm{OLS}}) = \tilde{\sigma}^2(\mathbf{X}^\top \mathbf{X})^{-1}$$

where $\tilde{\sigma}^2 = \sigma^2[\mathrm{Tr}(\mathbf{D}\mathbf{K}_Z) + 1]$

Unpacking $V(\hat{\gamma}^{\text{GLS}})$

$$
\begin{aligned}
\sigma^2(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1} &= \sigma^2\{\mathbf{X}^\top[\mathbf{Z}(\mathbf{D}\otimes\mathbf{I}_J)\mathbf{Z}^\top+\mathbf{I}]^{-1}\mathbf{X}\}^{-1}\\
&= \sigma^2\{\mathbf{X}^\top[\mathbf{I}-\mathbf{Z}(\mathbf{D}\otimes\mathbf{I}_J^{-1}+\mathbf{Z}^\top\mathbf{Z})\mathbf{Z}^\top]\mathbf{X}\}^{-1}\\
&= \sigma^2\{\mathbf{X}^\top\mathbf{X}-\mathbf{X}^\top\mathbf{Z}(\mathbf{D}\otimes\mathbf{I}_J^{-1}+\mathbf{Z}^\top\mathbf{Z})\mathbf{Z}^\top\mathbf{X}\}^{-1}\\
&= \sigma^2\{[(\mathbf{X}^\top\mathbf{X})^{-1}+(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Z}(\mathbf{D}\otimes\mathbf{I}_J)\mathbf{Z}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}]^{-1}\}^{-1}\\
&= \sigma^2[(\mathbf{X}^\top\mathbf{X})^{-1}+(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Z}(\mathbf{D}\otimes\mathbf{I}_J)\mathbf{Z}^\top\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}]
\end{aligned}
$$

We can substitute $\mathbf{X}$ with $\mathbf{X}^c$. As $\mathbf{Z}$ is block diagonal, ${\mathbf{X}^c}^\top\mathbf{Z}$ = $[{\mathbf{X}^c}_1^\top\mathbf{Z}_1|{\mathbf{X}^c}_2^\top\mathbf{Z}_2|\cdots|{\mathbf{X}^c}_J^\top\mathbf{Z}_J]$, and

$$
N\mathbf{G}={\mathbf{X}^c}^\top\mathbf{Z}(\mathbf{D}\otimes\mathbf{I}_J)\mathbf{Z}^\top\mathbf{X}^c=\sum_{j=1}^{J}{\mathbf{X}^c}_j^\top\mathbf{Z}_j\mathbf{D}\mathbf{Z}_j^\top{\mathbf{X}^c}_j
$$

is the cross-product matrix of ${\mathbf{X}^c}^\top\mathbf{Z}\mathbf{u}$, and $\mathbf{G}_{kl}$ seems equal to

$$
\begin{cases}
\tilde{n}N\mathbf{\Sigma}_{X_kZ}\mathbf{D}\mathbf{\Sigma}_{X_lZ}^\top & \text{when both } X_l \text{ and } X_k \text{ are purely level-1}\\
\tilde{n}N[\mathbf{\Sigma}_{X_kZ}\mathbf{D}\mathbf{\Sigma}_{X_lZ}^\top+\mathbf{K}_{X_kZ}\mathbf{D}\mathbf{K}_{X_lZ}^\top] & \text{otherwise}
\end{cases}
$$

where $\tilde{n}=\sum_{j=1}^J n_j^2/N$ (reduced to n if cluster sizes are equal across clusters). So

$$
\sigma^2({\mathbf{X}^c}^\top\mathbf{V}^{-1}\mathbf{X}^c)^{-}=\sigma^2[\mathbf{\Sigma}_X^{-1}+\tilde{n}\mathbf{\Sigma}_X^{-1}\mathbf{G}\mathbf{\Sigma}_X^{-1}],
$$

where $\mathbf{G}$ = $(\mathbf{\Sigma}_{XZ}\mathbf{D}\mathbf{\Sigma}_{XZ}^\top+\mathbf{F}\mathbf{K}_{XZ}\mathbf{D}\mathbf{K}_{XZ}^\top\mathbf{F})$, $\mathbf{F}$ is a diagonal filtering matrix in the form $\text{diag}[f_0, f_1, \ldots, f_{p-1}]$, where f_k = 1 if X_k has a level-2 component and 0 otherwise.

The OLS standard error and the MLM standard error have different rates of converagence to 0

For coefficient γ_k

$$\begin{aligned}
\mathrm{Deff}(\hat{\gamma}_k) &= \frac{\{\boldsymbol{\Sigma}_X^{-1}\}_{kk} + n\left[\boldsymbol{\Sigma}_X^{-1}\mathbf{G}\boldsymbol{\Sigma}_X^{-1}\right]_{kk}}{\{\boldsymbol{\Sigma}_X^{-1}\}_{kk}\left[\mathrm{Tr}(\mathbf{D}\mathbf{K}_Z) + 1\right]} \\
&= (1 - \mathrm{ICC}_D) + (1 - \mathrm{ICC}_D)n\frac{\left[\boldsymbol{\Sigma}_X^{-1}\mathbf{G}\boldsymbol{\Sigma}_X^{-1}\right]_{kk}}{\boldsymbol{\Sigma}_X^{-1}{}_{kk}}
\end{aligned}$$

- Generally speaking, deff is larger with a larger cluster size, n
- Deff is fixed for constant n (as in $\mathrm{Deff}[\hat{\mu}]$)

Example: Growth curve modeling with Tx $\times$ slope interaction

$$\mathbf{y}_i = \begin{bmatrix} 1 & T_i & 0 & 0 \\ 1 & T_i & 1 & T_i \\ 1 & T_i & 2 & 2T_i \\ 1 & T_i & 3 & 3T_i \\ 1 & T_i & 4 & 4T_i \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \\ e_{i4} \\ e_{i5} \end{bmatrix}$$

$$\boldsymbol{\gamma} = [1, 0, -0.1, -0.3]^\top, \mathbf{D} = \begin{bmatrix} 0.5 & \\ 0.2 & 0.1 \end{bmatrix}, \sigma^2 = 1.5$$

```
icc_d r2_GLMM
0.655   0.056
```

Example cont'd

```r
# Design effect
sigmaxz <- sigmax[, c(1, 3)]
Kxz <- Kx[, c(1, 3)]
F_mat <- diag(c(0, 1, 0, 1))  # filtering
G_mat <- (sigmaxz %*% D_mat %*% t(sigmaxz)) +
    F_mat %*% Kxz %*% D_mat %*% t(Kxz) %*% F_mat
sigmax_inv <- MASS::ginv(sigmax)
(1 - icc) * (1 + num_obs * diag(sigmax_inv %*% G_mat %*% sigmax_inv
    diag(sigmax_inv))
```

```
[1]       NaN 0.6321839 0.6896552 0.6896552
```

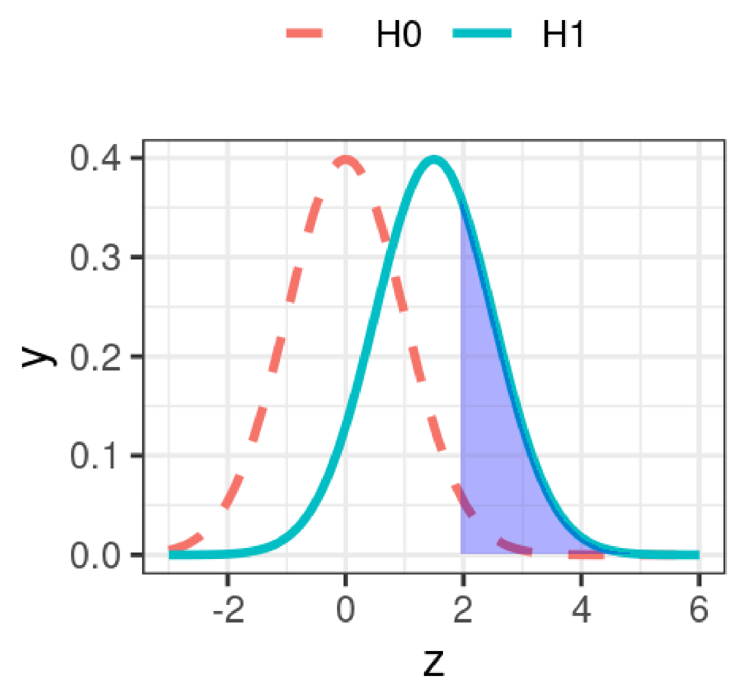
```r
m1 <- lmer(y ~ treat * time + (time | clus_id), data = dat)
m0 <- lm(y ~ treat * time, data = dat)
# Empirical variance ratio
diag(vcov(m1)) / diag(vcov(m0))
```

```
(Intercept)       treat        time  treat:time 
  0.8338815   0.8338815   0.6609629   0.6609629 
```

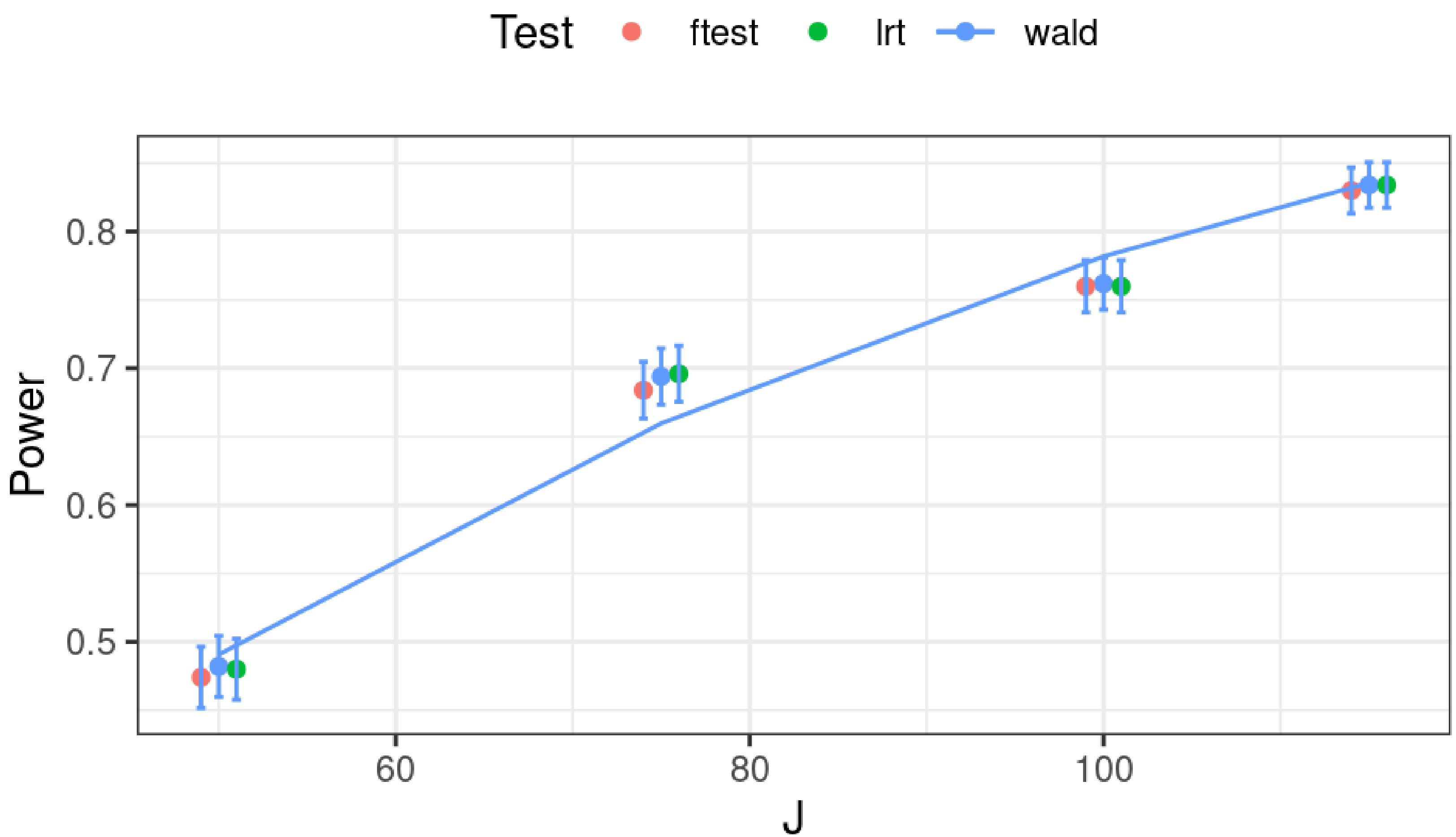
Power Analysis

For $H_0\colon \gamma_k = 0$

- Wald statistic: $z = \hat{\gamma}_k / \sqrt{\hat{V}(\hat{\gamma}_k)}$
  - In large sample, $z \sim N(z_1, 1)$, $z_1 = \gamma_k / \sqrt{V(\hat{\gamma}_k)}$
- With significance level α, power is approximately $\Phi(z_{1-\alpha/2} - z_1) + \Phi(z_{\alpha/2} - z_1)$



- Growth model example (Tx $\times$ time interaction)
  - Line: analytic formula; Points: 1,000 simulation samples



Summary

- Multicollinearity can happen between random-effect predictors, and between random- and fixed-effect predictors
- With random coefficients, one can do variance partitioning using the average variance of an observation
  - ICC_D quantifies the proportion of variance due to random effects
  - R^2 quantifies the proportion of variance due to fixed effects

Summary (cont'd)

- Standardized coefficients and standardized effect size can be defined
- Design effect quantifies variance inflation of an estimator due to cluster sampling
  - Can be applied to fixed-effect estimators
- Closed-form expression for power can be obtained using summary statistics

Thank You!

If you have any suggestions or feedback, please email me at hokchiol@usc.edu

References

Aguinis, H., & Culpepper, S. A. (2015). An expanded decision-making procedure for examining cross-level interaction effects With multilevel modeling. *Organizational Research Methods*, *18*(2), 155–176. https://doi.org/10.1177/1094428114563618

Hedges, L. V. (2009). Effect sizes in nested designs. In H. Cooper, L. V. Hedges, & J. C. Valentine, *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 337–355). Russell Sage Foundation.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944–946. https://doi.org/10.1111/2041-210X.12225

Lai, M. H. C. (2019). Correcting fixed effect standard errors when a crossed random effect was ignored for balanced and unbalanced designs. *Journal of Educational and Behavioral Statistics*, *44*(4), 448–472. https://doi.org/10.3102/1076998619843168

Lai, M. H. C., & Kwok, O. (2015). Examining the rule of thumb of not using multilevel modeling: The "design effect smaller than two" rule. *The Journal of Experimental Education*, *83*(3), 423–438. https://doi.org/10.1080/00220973.2014.907229

Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, *44*(2), 182–212. https://doi.org/10.1080/00273170902794214

Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*. https://doi.org/10.1037/met0000330

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184

Rights, J. D., & Sterba, S. K. (2023). R-squared measures for multilevel models with three or more levels. *Multivariate Behavioral Research*, *58*(2), 340–367. https://doi.org/10.1080/00273171.2021.1985948

Snijders, T. A. B., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, *18*(3), 237–259. https://doi.org/10.3102/10769986018003237

Supplemental Slides

Multicollinearity with Crossed Random Effects

$$\mathbf{y} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}_A\boldsymbol{\tau}_A + \mathbf{Z}_B\boldsymbol{\tau}_B + \epsilon$$

- $(\mathbf{M}_{X_0}\mathbf{Z}_A)^\top\mathbf{M}_{X_0}\mathbf{Z}_B$ = collinearity between the two crossed levels[1]
- When there are only random intercepts at Levels A and B,
  - For balanced designs (i.e., fully crossed), $\mathbf{Z}_A$ and $\mathbf{Z}_B$ are orthogonal (i.e., $\mathbf{M}_{X_0}\mathbf{Z}_A^\top\mathbf{M}_{X_0}\mathbf{Z}_B = \mathbf{0}$)
    - If Level B is omitted, its variance goes to Level 1
  - For unbalanced designs (i.e., partially crossed)
    - If Level B is omitted, some of its variance goes to Level A

```r
# Sample data of fully crossed random effects
library(lme4)
fm1 <- lmer(strength ~ (1 | batch) + (1 | cask),
Z <- getME(fm1, "Z")
Z0A <- Z[, 1:10]
Z0B <- Z[, 11:13]
crossprod(Z0B, Z0A)  # balanced design
```

```
3 x 10 sparse Matrix of class "dgCMatrix"

a 2 2 2 2 2 2 2 2 2 2
b 2 2 2 2 2 2 2 2 2 2
c 2 2 2 2 2 2 2 2 2 2
```

```r
# Sample data of fully crossed random effects
# Residual maker for grand intercept
MX0 <- diag(nrow(Z)) - matrix(1 / 60, nrow = nrow
# Cross-product is zero
crossprod(MX0 %*% Z0B, MX0 %*% Z0A) |> round(dig
```

```
3 x 10 Matrix of class "dgeMatrix"
  A B C D E F G H I J
a 0 0 0 0 0 0 0 0 0 0
b 0 0 0 0 0 0 0 0 0 0
c 0 0 0 0 0 0 0 0 0 0
```

See Lai (2019) and Luo & Kwok (2009) for more in-depth discussion

(Unconditional) ICC and R^2

From a random-intercepts model without fixed-effect predictors,

ICC = maximum R^2 for a level-2 predictor

1 - ICC = maximum R^2 for a purely level-1 predictor

(Conditional) ICC and R^2

From a conditional model with fixed-effect predictors,

ICC = Proportion of variance accounted for by level-2 random effects

R^2 = Proportion of variance accounted for by all fixed-effect predictors

SMD With Covariates

With covariates $\mathbf{X}_2, Y_{ij} = T_j\gamma_1 + \mathbf{x}_{2ij}^\top \boldsymbol{\gamma}_2 + Z_{0j}u_{0j} + e_{ij}$

- $\vec{T}^\top \mathbf{X}_2 = \mathbf{0}$ and no interactions

$$\delta = \frac{\gamma_1}{\sqrt{\boldsymbol{\gamma}_2^\top \boldsymbol{\Sigma}_{X_2} \boldsymbol{\gamma}_2 + \tau_0^2 + \sigma^2}}$$

With random slopes, $Y_{ij} = T_j\gamma_1 + \mathbf{x}_{2ij}^\top \boldsymbol{\gamma}_2 + \mathbf{z}_{ij}^\top \mathbf{u}_j + e_{ij}$

- and also $\vec{T}^\top \mathbf{Z} = \mathbf{0}$

$$\delta = \frac{\gamma_1}{\sqrt{\boldsymbol{\gamma}_2^\top \boldsymbol{\Sigma}_{X_2} \boldsymbol{\gamma}_2 + \sigma^2 [\mathrm{Tr}(\mathbf{D}\mathbf{K}_Z) + 1]}}$$

Code example (data)

y	treat	x2	clus_id
<dbl>	<dbl>	<int>	<int>
2.7426168	0	0	1
2.9176709	0	1	1
1.2855431	0	1	1
4.1538467	0	1	1
2.6109176	0	1	1
4.2501591	0	1	2
5.3785220	0	0	2
4.6200485	0	0	2
4.4757641	0	1	2
5.0979586	0	1	2

1-10 of 150 rows

Previous **1** 2 3 4 5 6 ... 15 Next

Code example (cont'd)

```
library(lme4)
# Fit MLM
m1 <- lmer(y ~ treat + x2 + (x2 | clus_id), data = sim_data)
# Variance by x2
Sigma_x <- with(sim_data, { x2c <- x2 - mean(x2); mean(x2c^2) })
va_x2 <- Sigma_x * fixef(m1)[["x2"]]^2
# Variance by random intercepts and random slopes
# Using parameterization in lme4
Kz <- crossprod(cbind(1, sim_data$x2)) / nrow(sim_data)
va_z <- sum(VarCorr(m1)$clus_id * Kz)  # Tr(Tau, Kz)
# Show all components
c(gamma_1 = fixef(m1)[["treat"]], va_x2 = va_x2,
  va_z = va_z, va_e = sigma(m1)^2)
```

```
    gamma_1       va_x2        va_z        va_e 
0.257714138 0.001712517 0.537828861 0.989420368 
```

```
# Standardized effect size
fixef(m1)[["treat"]] / sqrt(va_x2 + va_z + sigma(m1)^2)
```

```
[1] 0.2084203
```

Additional Notes on SMD

- *SE* and CI: Delta method, bootstrap, Bayesian (also approximate variance as in Hedges, 2009)
- Using unadjusted (marginal) mean difference when $\vec{T}^\top \mathbf{X}_2 \neq \mathbf{0}$
  - Unadjusted mean difference: $E(Y|T=1,\mathbf{u})$ - $E(Y|T=1,\mathbf{u})$ = $\gamma_1 + \Sigma_{TT}^{-1}\mathbf{\Sigma}_{TX_2}\boldsymbol{\gamma}_2{}^1$
  - Unadjusted within-arm variance by fixed effects: $V(\hat{Y}|T,\mathbf{u}) = \boldsymbol{\gamma}^\top\mathbf{\Sigma}_X\boldsymbol{\gamma} - \gamma_1^2\Sigma_{TT} - 2\gamma_1\mathbf{\Sigma}_{TX_2}\boldsymbol{\gamma}_2^\top - \Sigma_{TT}^{-1}\boldsymbol{\gamma}_2^\top\mathbf{\Sigma}_{X_2T}\mathbf{\Sigma}_{TX_2}$
  - More complicated when T also correlates with $\mathbf{Z}$

Choice of Denominator in SMD

- Square root of total variance should be default[1]
- Longitudinal: between-person variance
  - i.e., variance at time 0 (or time t?)
- Multisite trials: Exclude treatment effect heterogeneity (i.e., random slope of T)?
  - i.e., total variance without the intervention

Power Analysis

Wait, This Is Not New . . .

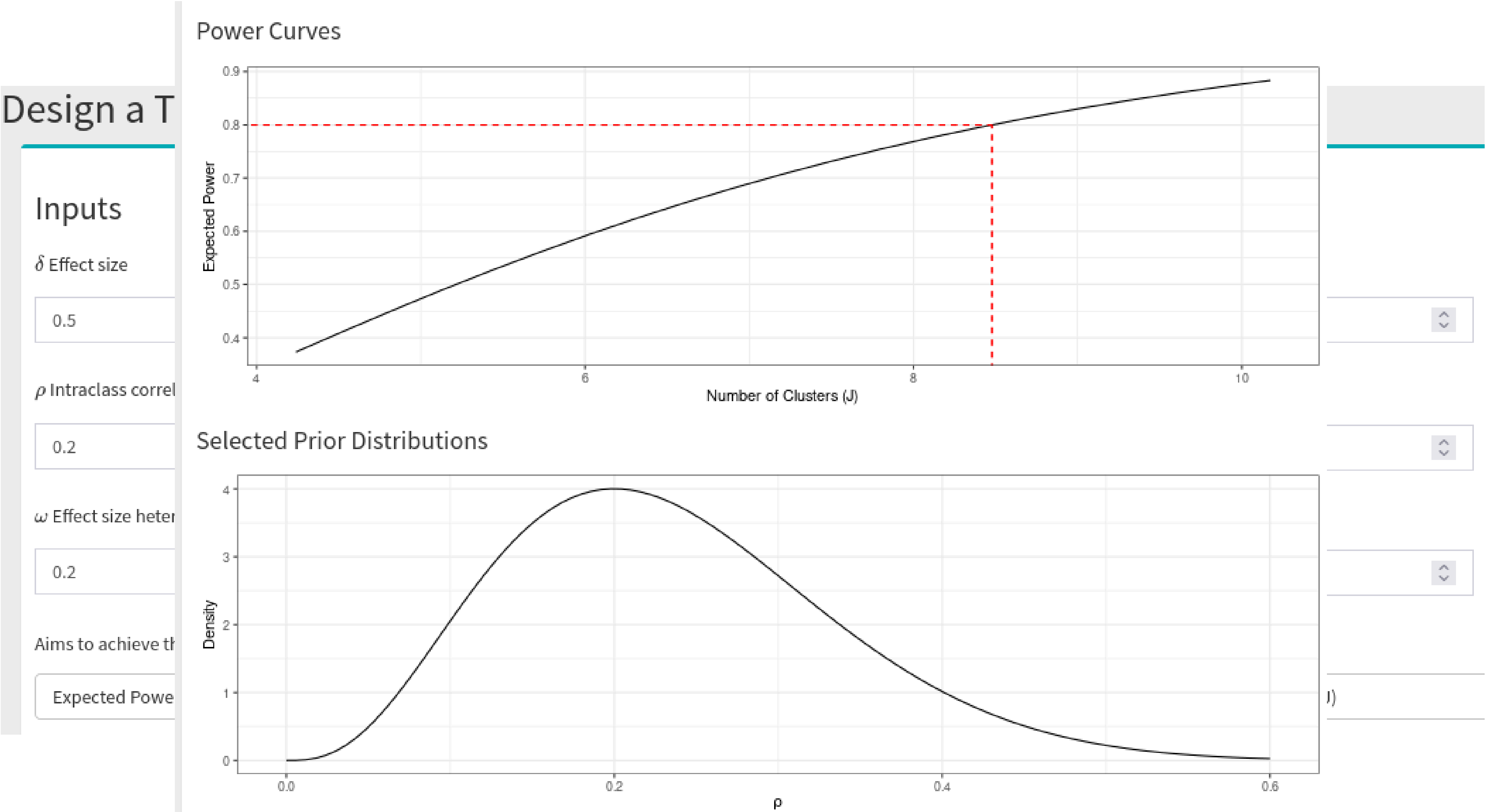
- Snijders & Bosker (1993) also had closed-form expressions for $V(\hat{\gamma}^{\mathrm{GLS}})$
  - Implemented in the PINT software[1]
- Murayama et al. (2022): summary-statistics-based power analysis[2]
  - Primarily useful when prior studies reported relevant t statistic

What can still be done?

- Implementing formula-based approach in other software
- Express $V(\hat{\gamma}^{\mathrm{GLS}})$ in terms of effect sizes
- Extend to models with other error covariance structure and more levels
- Handling uncertainty in nuisance parameters (e.g., ICC)[1]

Handling Nuisance Parameters in Power Analysis

- There are many nuisance parameters (e.g., ICC, correlation among predictors) that affect power
  - Usually we just choose some convenient values
- A more disciplined approach: incorporate our uncertainty of those parameters as Bayesian priors
  - See the HCB shiny application by Winnie Tse



Shiny app: https://winnie-wy-tse.shinyapps.io/hcb_shiny/

R package *hcbr*: https://github.com/winniewytse/hcbr