Outline

- Inference for fixed effects.
- Inference for variance components.
- Global measures of fit.
- Computer Lab 3

Reading: Snijders & Bosker, Chapter 6
Additional References

Inference for Fixed Effects

**Goal:** Make inferences about model parameters and make generalizations from a specific sample to the population from which the sample was selected.

- Approximate Wald tests ($z$ tests).
- Approximate $t$ and $F$ tests.
- Robust estimation.
- Likelihood ratio tests.
Approximate Wald Tests

Need the sampling distribution of the fixed parameter estimates, \( \hat{\Gamma} = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots)' \).

The asymptotic sampling distribution of \( \hat{\Gamma} \) is

\[
\hat{\Gamma} \sim N \left( \Gamma, \text{cov}(\hat{\Gamma}) \right)
\]

where

\[
\hat{\Gamma} = \left( \sum_{j=1}^{N} x_j' \hat{V}_j^{-1} x_j \right)^{-1} \sum_{j=1}^{N} x_j' \hat{V}_j^{-1} y_j
\]

where \( \hat{V}_j \) is the estimated covariance matrix of \( Y_j \), which equals

\[
\hat{V}_j = Z_j \hat{T} Z_j' + \hat{\sigma}^2 I
\]

Our estimate of \( \Gamma \) depends on \( \hat{T} \) and \( \hat{\sigma}^2 \).
Covariance Matrix of $\hat{\Gamma}$

To get an estimate of $\hat{\Gamma}$:

**IF**

- The model for the mean of $Y_j$ is correctly specified, (i.e., $X_j \Gamma$) so $E(\hat{\Gamma}) = \Gamma$ (i.e, unbiased).

- The marginal covariance matrix is correctly specified, (i.e., $V_j = Z_j T Z_j' + \sigma^2 I$) so the covariance matrix of data equals the predicted covariance matrix.
**I** Covariance Matrix of $\hat{\Gamma}$

THEN

$$\widehat{\text{cov}}(\hat{\Gamma}) = \left( \sum_{j=1}^{N} x_j' \hat{V}_j^{-1} x_j \right)^{-1} = (X' \hat{V}^{-1} X)^{-1}$$

where

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} V_1 & 0 & \ldots & 0 \\ 0 & V_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & V_N \end{pmatrix}$$

We can now use the fact $\hat{\Gamma} \sim \mathcal{N} \left( \Gamma, \text{cov}(\hat{\Gamma}) \right)$. 
**Digression: Distribution of $\hat{\Gamma}$**

- Since $Y \sim N(X\Gamma, \Sigma_Y)$ and

\[
\hat{\Gamma} = \left( \sum_{j=1}^{N} x_j' \hat{\Sigma}_j^{-1} x_j \right)^{-1} \sum_{j=1}^{N} x_j' \hat{\Sigma}_j^{-1} y_j
\]

\[
= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y
\]

\[
= A\ y
\]

- The Expected value,

\[
E(\hat{\Gamma}) = E[(X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y]
\]

\[
= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}E[(X\Gamma + \epsilon)]
\]

\[
= (X'\hat{\Sigma}^{-1}X)^{-1}(X'\hat{\Sigma}^{-1}X)E[(\Gamma + \epsilon)]
\]

\[
= \Gamma
\]
I. Distribution of $\hat{\Gamma}$ (continued)

- Covariance matrix,

$$
\text{cov}(\hat{\Gamma}) = AVA'
$$

$$
= [(X'V^{-1}X)^{-1}X'V^{-1}] \sum_Y [V^{-1}X(X'V^{-1}X)^{-1}]
$$

$$
= (X'V^{-1}X)^{-1}X'V^{-1}X(X'V^{-1}X)^{-1}
$$

$$
= (X'V^{-1}X)^{-1}
$$

- Since $\hat{\Gamma}$ is a linear function of a vector of normal random variables (i.e., $Y$), $\hat{\Gamma}$ is normal.

- So $\hat{\Gamma} \sim \mathcal{N}(\Gamma, (X'V^{-1}X)^{-1})$
Approximate Wald Tests

- Perform statistical hypothesis tests on
  - One $\gamma$, e.g.,
    $$H_0 : \gamma_0 = 0 \text{ versus } H_o : \gamma_0 \neq 0$$
  - Multiple $\gamma$’s, including contrasts, e.g.,
    $$H_0 : L\Gamma = 0 \text{ versus } H_a : L\Gamma \neq 0$$
- Form confidence intervals for parameters.
One Fixed Effect

**Sampling distribution** for one fixed effect,

\[ \hat{\gamma}_{kl} \sim N(\gamma_{kl}, \text{var}(\hat{\gamma}_{kl})) \]

**Statistical Hypothesis:**

\[ H_o: \gamma_{kl} = \gamma^*_{kl} \quad \text{versus} \quad H_a: \gamma_{kl} \neq \gamma^*_{kl}. \]

**Note:**

- Usually, \( \gamma^*_{kl} = 0 \)
- Can do directional tests, i.e.,
  \[ H_a: \gamma_{kl} > \gamma^*_{kl} \quad \text{or} \quad H_a: \gamma_{kl} < \gamma^*_{kl} \]

Test statistic and approximate sampling distribution:

\[ z = \frac{\hat{\gamma}_{kl} - \gamma^*_{kl}}{SE} \sim N(0,1) \quad \text{or} \quad z^2 \sim \chi^2_1 \]
Wald Test: Example

HSB — a really complex model

**Level 1:**

\[(\text{math})_{ij} = \beta_0j + \beta_{1j}(\text{cSES})_{ij} + \beta_{2j}(\text{female})_{ij} + \beta_{3j}(\text{minority})_{ij} + R_{ij}\]

**Level 2:**

\[
\begin{align*}
\beta_0j &= \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\text{SES})_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{sector})_j + \gamma_{12}(\text{size})_j + U_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}(\text{sector})_j + \gamma_{22}(\text{size})_j + \gamma_{23}(\text{SES})_j + U_{2j} \\
\beta_{3j} &= \gamma_{30} + \gamma_{31}(\text{sector})_j + \gamma_{32}(\text{size})_j + \gamma_{33}(\text{SES})_j + U_{3j}
\end{align*}
\]
HSB: Linear Mixed Model

\[(\text{math})_{ij} = [\gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\text{SES})_j] + [\gamma_{10} + \gamma_{11}(\text{sector})_j + \gamma_{12}(\text{size})_j](\text{cSES})_{ij} + [\gamma_{20} + \gamma_{21}(\text{sector})_j + \gamma_{22}(\text{size})_j + \gamma_{23}(\text{SES})_j](\text{female})_{ij} + [\gamma_{30} + \gamma_{31}(\text{sector})_j + \gamma_{32}(\text{size})_j + \gamma_{33}(\text{SES})_j](\text{minority})_{ij} + U_{0j} + U_{ij}(\text{cSES})_{ij} + U_{2j}(\text{female})_{ij} + U_{3j}(\text{minority})_{ij} + R_{ij} \]

\[= \gamma_{00} + \gamma_{10}(\text{cSES})_{ij} + \gamma_{20}(\text{female})_{ij} + \gamma_{30}(\text{minority})_{ij} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\text{SES})_j + \gamma_{11}(\text{sector})_j(\text{cSES})_{ij} + \gamma_{12}(\text{size})_j(\text{cSES})_{ij} + \gamma_{21}(\text{sector})_j(\text{female})_{ij} + \gamma_{22}(\text{size})_j(\text{female})_{ij} + \gamma_{23}(\text{SES})_j(\text{female})_{ij} + \gamma_{31}(\text{sector})_j(\text{minority})_{ij} + \gamma_{32}(\text{size})_j(\text{minority})_{ij} + \gamma_{33}(\text{SES})_j(\text{minority})_{ij} + U_{0j} + U_{1j}(\text{cSES})_{ij} + U_{2j}(\text{female})_{ij} + U_{3j}(\text{minority})_{ij} + R_{ij} \]
**HSB: SAS/MIXED Input**

**PROC MIXED** data=hsbcent noclprint covtest method=ML ic;
**CLASS** id;
**MODEL** mathach = cSES female minority meanSES size sector
cSES*size cSES*sector female*meanSES female*size
female*sector minority*meanSES minority*size minority*sector
/solution chisq;
**RANDOM** intercept female minority cSES / subject=id type=un;
**RUN;**
SAS/MIXED & Wald Tests

To get the Wald test statistic and \( p \)-values, you need to specify the “chisq” option in the model statement.

The null hypothesis is

\[ H_0 : \gamma_{kl} = 0 \quad \text{versus} \quad H_a : \gamma_{kl} \neq 0. \]

It gives you “chi-square” (i.e., \( z^2 \)), so if you want to do a one-tailed test or use a different value in the null hypothesis, you need to compute \( z \) by hand.
### Solution for Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Estimate</th>
<th>se</th>
<th>DF</th>
<th>Wald</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.0260</td>
<td>0.4691</td>
<td>1</td>
<td>53.81</td>
<td>&lt; .0001 *</td>
</tr>
<tr>
<td>cses</td>
<td>2.2823</td>
<td>0.3111</td>
<td>1</td>
<td>53.81</td>
<td>&lt; .0001 *</td>
</tr>
<tr>
<td>female</td>
<td>−0.3402</td>
<td>0.4637</td>
<td>1</td>
<td>0.54</td>
<td>0.4632</td>
</tr>
<tr>
<td>minority</td>
<td>−4.2068</td>
<td>0.6714</td>
<td>1</td>
<td>39.26</td>
<td>&lt; .0001 *</td>
</tr>
<tr>
<td>meanses</td>
<td>4.2207</td>
<td>0.5003</td>
<td>1</td>
<td>71.17</td>
<td>&lt; .0001 *</td>
</tr>
<tr>
<td>size</td>
<td>0.001125</td>
<td>0.000314</td>
<td>1</td>
<td>12.86</td>
<td>0.0003 *</td>
</tr>
<tr>
<td>sector</td>
<td>1.7360</td>
<td>0.4173</td>
<td>1</td>
<td>17.31</td>
<td>&lt; .0001 *</td>
</tr>
<tr>
<td>cses*size</td>
<td>0.000032</td>
<td>0.000203</td>
<td>1</td>
<td>0.03</td>
<td>0.8728</td>
</tr>
<tr>
<td>cses*sector</td>
<td>−1.0033</td>
<td>0.2528</td>
<td>1</td>
<td>15.74</td>
<td>&lt; .0001 *</td>
</tr>
<tr>
<td>female*meanses</td>
<td>−0.03207</td>
<td>0.4838</td>
<td>1</td>
<td>0.00</td>
<td>0.9471</td>
</tr>
<tr>
<td>female*size</td>
<td>−0.00070</td>
<td>0.000304</td>
<td>1</td>
<td>5.36</td>
<td>0.0206 *</td>
</tr>
<tr>
<td>female*sector</td>
<td>−0.3006</td>
<td>0.4284</td>
<td>1</td>
<td>0.49</td>
<td>0.4829</td>
</tr>
<tr>
<td>minority*meanses</td>
<td>−0.7793</td>
<td>0.5391</td>
<td>1</td>
<td>2.09</td>
<td>0.1483</td>
</tr>
<tr>
<td>minority*size</td>
<td>0.000183</td>
<td>0.000398</td>
<td>1</td>
<td>0.21</td>
<td>0.6446</td>
</tr>
<tr>
<td>minority*sector</td>
<td>2.1189</td>
<td>0.5430</td>
<td>1</td>
<td>15.23</td>
<td>&lt; .0001 *</td>
</tr>
</tbody>
</table>
Confidence Intervals for $\gamma_{kl}$’s

Given the estimated standard errors and fixed effects, we can construct $(1 - \alpha)100\%$ confidence intervals for $\gamma_{kl}$’s:

$$\hat{\gamma}_{kl} \pm z_{\alpha/2} \hat{SE}$$

For example, a 95% confidence interval for $\gamma_{10}$, the coefficient for $(cSES)_{ij}$, is

$$2.2823 \pm 1.96(0.3111) \rightarrow (1.67, 2.89)$$
We may want to

- Simultaneously test a set of \( \gamma \)'s.
  - Consider whether to drop multiple effects from the model all at once.
  - For discrete variables where you've entered effect or dummy codes for the levels of the variable (rather than using the `CLASS` statement and letting SAS create dummy codes).

- One or more contrasts of \( \gamma \)'s (e.g., to test whether some \( \gamma \)'s are equal).
General Tests on Fixed Effects

For the general case, tests are based on the fact

\[ \hat{\Gamma} \sim \mathcal{N}(\Gamma, \text{cov}(\hat{\Gamma})) \]

Hypotheses are in the form of

\[ H_0 : L\Gamma = 0 \quad \text{versus} \quad H_a : L\Gamma \neq 0 \]

where \( L \) is an \((c \times p)\) matrix of constants that define the hypothesis tests.

In Scaler From:

\[ H_{o(1)} : \sum_{k=1}^{p} l_{1k}\gamma_k = 0, \quad H_{o(2)} : \sum_{k=1}^{p} l_{2k}\gamma_k = 0, \quad \ldots \quad H_{o(c)} \]

- \( l_{rk} \) = a constant in the \( r^{th} \) row and \( k^{th} \) column of matrix \( L \).
- \( c \) = number of hypothesis tests (rows of \( L \)).
- \( p \) = number of parameters for fixed effects (elements in \( \Gamma \)).
- \( c \leq p \).
General Test Statistic

\[
\hat{\Gamma}'L'(L \left( \sum_{j=1}^{N} X_j' \hat{V}_j^{-1} X_j \right)^{-1} L')^{-1} L\hat{\Gamma}
\]

covariance matrix of \( L\hat{\Gamma} \)

and asymptotically follows a \( \chi^2 \) distribution with \( df = c \), the number of rows in \( L \) (i.e., the rank of \( L \)).

I won’t make you compute this by hand... Let SAS do the busy-work.
In our example $\Gamma$ is a $(15 \times 1)$ vector:

$$\Gamma' = (\gamma_{00}, \gamma_{10}, \gamma_{20}, \gamma_{30}, \gamma_{01}, \gamma_{02}, \gamma_{03}, \gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}, \gamma_{23}, \gamma_{31}, \gamma_{32}, \gamma_{33})$$

From the Wald tests, we found that the following cross-level interactions were not significant:

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Parameter</th>
<th>Interaction</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(\text{size})<em>j(\text{cSES})</em>{ij}$</td>
<td>$\gamma_{12}$</td>
<td>$(\text{size})<em>j(\text{minority})</em>{ij}$</td>
<td>$\gamma_{32}$</td>
</tr>
<tr>
<td>$(\text{sector})<em>j(\text{female})</em>{ij}$</td>
<td>$\gamma_{21}$</td>
<td>$(\text{SES})<em>j(\text{minority})</em>{ij}$</td>
<td>$\gamma_{33}$</td>
</tr>
<tr>
<td>$(\text{SES})<em>j(\text{female})</em>{ij}$</td>
<td>$\gamma_{23}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simultaneously Testing $\gamma_{rk}$'s

We can simultaneously test all of these cross-level interactions by defining $(5 \times 15)$ matrix,

$$
\Gamma' = \begin{pmatrix}
\gamma_{00} & \gamma_{10} & \gamma_{20} & \gamma_{30} & \gamma_{01} & \gamma_{02} & \gamma_{03} & \gamma_{11} & \gamma_{12} & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{31} & \gamma_{32} & \gamma_{33}
\end{pmatrix}
$$

$$
L = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
$$
Simultaneously Testing Cross-Level

Statistical hypotheses are

\[ H_0 : \mathbf{L}\Gamma = \begin{pmatrix} \gamma_{12} \\ \gamma_{21} \\ \gamma_{23} \\ \gamma_{32} \\ \gamma_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{vs} \quad H_a : \mathbf{L}\Gamma \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \]
**SAS/MIXED for Simultaneous Tests**

CONTRAST 'Cross-level interactions'

- cSES*size 1
- female*meanSES 1
- female*sector 1
- minority*meanSES 1
- minority*size 1 / chisq;

- “CONTRAST” statement specifies the effect that you want to test.
- We only need to enter a single value because each of these interactions has only a single parameter estimated.

---

**SAS/MIXED Output:**

<table>
<thead>
<tr>
<th>Contrasts</th>
<th>DF</th>
<th>Chi-Square</th>
<th>Pr&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-level interactions</td>
<td>5</td>
<td>2.97</td>
<td>0.7048</td>
</tr>
</tbody>
</table>
SAS/MIXED Contrast Statement

- If a variable has 5 levels. For example, hours watching TV from the TIMSS data set used in lab where it is entered as a nominal variable. To test whether levels the differences between levels 1, 2, 3, and 4 are different:

  ```
  CONTRAST 'Any differences between levels 1 to 4?'
    hours_computer_games 1 -1 0 0 0,
    hours_computer_games 1 0 -1 0 0,
    hours_computer_games 1 0 0 -1 0;
  ```

- If you want to test whether the average of 1–4 is different from level 5:

  ```
  CONTRAST 'Level 1–4 versus level 5'
    hours_computer_games 1 1 1 1 -4;
  ```

- You can have multiple contrast statements.
SAS/MIXED Input

- The `CONTRAST` statement only gives the test statistic, $df$ and $p$-value.

- The `ESTIMATE` statement is exactly like `CONTRAST`, except
  - Can only enter 1 row of $L$.
  - Output includes $L\hat{\Gamma}$ and it’s the S.E. of $L\hat{\Gamma}$, as well as the $df$, test statistics and $p$-value.
Contrasts in R

- I couldn’t figure out how to do them in R (at least like what’s on previous pages so I wrote a function)
- Run the function “contrast.txt” in the console.
- Create $L$ that has rows are tests/constrasts and columns correspond to fixed effects.
- `contrast(model,L)`
- Returns table with F, numerator df, a guess at denominator df, Wald $X^2$, df, and p-value for Wald. At a later date, I will add options for denominator df for the F test.
Contrasts in R

```r
require(lmerTest)
cmodel <- lmer(mathach 1 + cses + female + minority
+ meanses + sdsize + sector + cses*sdsize +
cses*sector + female*meanses + female*sdsize +
female*sector + minority*meanses + minority*sdsize +
minority*sector + (1 + cses + female | id), data=hsb,
REML=FALSE)
L <- matrix(0,nrow=5,ncol=15)
L[5,14] <- 1
L[4,13] <- 1
L[3,11] <- 1
L[2,10] <- 1
L[1, 8] <- 1
contrast(cmodel,L)
```

<table>
<thead>
<tr>
<th>F</th>
<th>num df</th>
<th>den df</th>
<th>p-value</th>
<th>X2</th>
<th>df</th>
<th>p-chisquare</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.73286</td>
<td>5.00000</td>
<td>156.15317</td>
<td>0.13014</td>
<td>8.66430</td>
<td>5.00000</td>
<td>0.123</td>
</tr>
</tbody>
</table>
Problem With Wald Tests

The estimated standard errors used in the Wald tests do not take into account the variability introduced by estimating the variance components.

The estimated standard errors are too small $\rightarrow$ Wald tests are a bit too “liberal” (i.e., the $p$–values are too small).

Solution: Use approximate $t$– and $F$– statistics.
Approximate $t$-tests and $F$-tests

For hypothesis tests and/or confidence intervals for a single $\gamma$, use Students $t$–distribution instead of the standard normal.

The test statistic still equals

$$
\frac{\hat{\gamma}_{kl} - \gamma^*_l}{SE}
$$

But it is compared to a $t$–distribution where the degrees of freedom are estimated from the data.
### Example: Approximate $t$-tests

| Effect          | Estimate | se    | DF  | $t$   | Pr>|$t$| |
|-----------------|----------|-------|-----|-------|-----|-----|
| Intercept       | 12.0260  | 0.4691| 155 | 25.63 | < .0001 |
| cses            | 2.2823   | 0.3111| 157 | 7.34  | < .0001 |
| female          | -0.3402  | 0.4637| 121 | -0.73 | 0.4646 |
| minority        | -4.2068  | 0.6714| 133 | -6.27 | < .0001 |
| meanses         | 4.2207   | 0.5003| 6604| 8.44  | < .0001 |
| size            | 0.001125 | 0.000314| 6604| 3.59  | 0.0003 |
| sector          | 1.7360   | 0.4173| 6604| 4.16  | < .0001 |
| cses*size       | 0.000032 | 0.000203| 6604| 0.16  | 0.8729 |
| cses*sector     | -1.0033  | 0.2528| 6604| -3.97 | < .0001 |
| female*meanses  | -0.03207 | 0.4838| 6604| -0.07 | 0.9471 |
| female*size     | -0.00070 | 0.000304| 6604| -2.31 | 0.0207 |
| female*sector   | -0.3006  | 0.4284| 6604| -0.70 | 0.4829 |
| minority*meanses| -0.7793 | 0.5391| 6604| -1.45 | 0.1484 |
| minority*size   | 0.000183 | 0.000398| 6604| 0.46  | 0.6446 |
| minority*sector | 2.1189   | 0.5430| 6604| 3.90  | < .0001 |
Approximate $F$-tests

For multiple tests and/or contrasts performed simultaneously, use the $F$-statistic

$$F = \frac{\hat{\Gamma}' L' \left[ L \left( \sum_{j=1}^{N} X_j' \hat{V}_j^{-1} X_j \right)^{-1} L' \right]^{-1} L \hat{\Gamma}}{c}$$

which is compared to an $F$ distribution where the numerator degrees of freedom equals $c$ (i.e., rank of $L$, number of tests/contrasts performed; that is, the number of rows in $L$). The denominator $df$ are estimated from the data.
Degrees of Freedom

There are 6 options in SAS/MIXED for determining the degrees of freedom which will be used in tests for fixed effects produced by MODEL, CONTRAST and ESTIMATE statements (and LSMEANS, which we haven’t talked about).

The options are:

- `ddf= value`. You specify your own value.

- `ddfm=contain`. This is the “containment” method and it is the default when you have a RANDOM statement.
 Degrees of Freedom (continued)

- ddfm=residual. This equals $n_+ - \text{(number of parameters estimated)}$.

- ddfm=betwithin. This is the default when you have a REPEATED statement and recommended instead of contain when the $Z_j$ matrices have a large number of columns.

  - The residual degrees of freedom are divided into a between-group and within-group part.
  - If the fixed effect changes within a group, $df$ is set equal to the within-group portion.
  - If the fixed effect does not change within a group (i.e., a macro level variable), SAS sets $df$ equal to the between-group portion.
Degrees of Freedom (continued)

- **ddfm=satterth.** General Satterthwaite approximation; based on the data. Works well with moderate to large samples; small sample properties unknown.

- **ddfm=kenwardroger.** Based on the data. It adjusts estimated covariance matrix for the fixed and random effects and then computes Satterthwaite approximation.
Simulated: \( N = 160, \, n_j = 10, \, n_+ = 1600, \) & \( p = 3 \)

| ddfm= Effect | Effect | Estimate | s.e. | DF | \( t \) | Pr > \(| t |\) |
|---|---|---|---|---|---|---|
| Contain Intercept | 12.0368 | .2753 | 158 | 43.72 | < .01 |
| \( x \) | 1.9930 | .1584 | 1439 | 12.58 | < .01 |
| \( z \) | 3.1423 | .2804 | 1439 | 11.20 | < .01 |
| Residual Intercept | 12.0368 | .2753 | 1597 | 43.72 | < .01 |
| \( x \) | 1.9930 | .1584 | 1597 | 12.58 | < .01 |
| \( z \) | 3.1423 | .2804 | 1597 | 11.20 | < .01 |
| Betwithin Intercept | 12.0368 | .2753 | 158 | 43.72 | < .01 |
| \( x \) | 1.9930 | .1584 | 1439 | 12.58 | < .01 |
| \( z \) | 3.1423 | .2804 | 158 | 11.20 | < .01 |
| Satterh Intercept | 12.0368 | .2753 | 160 | 43.72 | < .01 |
| \( x \) | 1.9930 | .1584 | 1543 | 12.58 | < .01 |
| \( z \) | 3.1423 | .2804 | 160 | 11.20 | < .00 |
| Kenward-Rogers Intercept | 12.0368 | .2753 | 160 | 43.72 | < .01 |
| \( x \) | 1.9930 | .1585 | 1543 | 12.58 | < .01 |
| \( z \) | 3.1423 | .2804 | 160 | 11.20 | < .01 |
Example: SAS Input for HSB

PROC MIXED data=hsbcent noclprint covtest method=ML;

CLASS id;

MODEL mathach = cSES female minority meanSES size sector 
cSES*size cSES*sector female*meanSES female*size 
female*sector minority*meanSES minority*size minority*sector 
/ solution chisq ddfM=satterth cl alpha=.01 ;

RANDOM intercept female minority cSES / subject=id type=un;

CONTRAST 'Cross-level interactions'
  cSES*size 1, 
  female*meanSES 1, 
  female*sector 1, 
  minority*meanSES 1, 
  minority*size 1 / chisq ddfm=satterth;
Output: Model Information

Data Set: WORK.HSBCENT
Dependent Variable: mathach
Covariance Structure: Unstructured
Subject Effect: id
Estimation Method: ML
Residual Variance Method: Profile
Fixed Effects SE Method: Model-Based
Degrees of Freedom Method: Satterthwaite
## Output: Solution for Fixed Effects

| Effect          | Estimate | se   | DF  | t     | Pr > |t||
|-----------------|----------|------|-----|-------|------|----|
| Intercept       | 12.0260  | 0.4691 | 139 | 25.63 | < .0001 |
| cses            | 2.2823   | 0.3111 | 148 | 7.34  | < .0001 |
| female          | −0.3402  | 0.4637 | 123 | −0.73 | 0.4646 |
| minority        | −4.2068  | 0.6714 | 157 | −6.27 | < .0001 |
| meanses         | 4.2207   | 0.5003 | 182 | 8.44  | < .0001 |
| size            | 0.001125 | 0.000314 | 156 | 3.59  | 0.0004 |
| sector          | 1.7360   | 0.4173 | 134 | 4.16  | < .0001 |
| cses*size       | 0.000032 | 0.000203 | 155 | 0.16  | 0.8731 |
| cses*sector     | −1.0033  | 0.2528 | 148 | −3.97 | 0.0001 |
| female*meanses  | −0.03207 | 0.4838 | 166 | −0.07 | 0.9472 |
| female*size     | −0.00070 | 0.000304 | 133 | −2.31 | 0.0222 |
| female*sector   | −0.3006  | 0.4284 | 143 | −0.70 | 0.4840 |
| minority*meanses| −0.7793  | 0.5391 | 120 | −1.45 | 0.1509 |
| minority*size   | 0.000183 | 0.000398 | 142 | 0.46  | 0.6453 |
| minority*sector | 2.1189   | 0.5430 | 133 | 3.90  | 0.0002 |
Output: Type 3 Tests of Fixed Effects

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Chi-Square</th>
<th>F Value</th>
<th>Pr &gt; ChiSq</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>cses</td>
<td>1</td>
<td>148</td>
<td>53.81</td>
<td>53.81</td>
<td>&lt; .0001</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>female</td>
<td>1</td>
<td>123</td>
<td>0.54</td>
<td>0.54</td>
<td>0.4632</td>
<td>0.4646</td>
</tr>
<tr>
<td>minority</td>
<td>1</td>
<td>157</td>
<td>39.26</td>
<td>39.26</td>
<td>&lt; .0001</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>meanses</td>
<td>1</td>
<td>182</td>
<td>71.17</td>
<td>71.17</td>
<td>&lt; .0001</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>size</td>
<td>1</td>
<td>156</td>
<td>12.86</td>
<td>12.86</td>
<td>0.0003</td>
<td>0.0004</td>
</tr>
<tr>
<td>sector</td>
<td>1</td>
<td>134</td>
<td>17.31</td>
<td>17.31</td>
<td>&lt; .0001</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>cses*size</td>
<td>1</td>
<td>155</td>
<td>0.03</td>
<td>0.03</td>
<td>0.8728</td>
<td>0.8731</td>
</tr>
<tr>
<td>cses*sector</td>
<td>1</td>
<td>148</td>
<td>15.74</td>
<td>15.74</td>
<td>&lt; .0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>female*meanses</td>
<td>1</td>
<td>166</td>
<td>0.00</td>
<td>0.00</td>
<td>0.9471</td>
<td>0.9472</td>
</tr>
<tr>
<td>female*size</td>
<td>1</td>
<td>133</td>
<td>5.36</td>
<td>5.36</td>
<td>0.0206</td>
<td>0.0222</td>
</tr>
<tr>
<td>female*sector</td>
<td>1</td>
<td>143</td>
<td>0.49</td>
<td>0.49</td>
<td>0.4829</td>
<td>0.4840</td>
</tr>
<tr>
<td>minority*meanses</td>
<td>1</td>
<td>120</td>
<td>2.09</td>
<td>2.09</td>
<td>0.1483</td>
<td>0.1509</td>
</tr>
<tr>
<td>minority*size</td>
<td>1</td>
<td>142</td>
<td>0.21</td>
<td>0.21</td>
<td>0.6446</td>
<td>0.6453</td>
</tr>
<tr>
<td>minority*sector</td>
<td>1</td>
<td>133</td>
<td>15.23</td>
<td>15.23</td>
<td>&lt; .0001</td>
<td>0.0002</td>
</tr>
</tbody>
</table>
## Output: 99% Confidence Limits

Produced by the “`cl alpha=.01`” option in the `MODEL` statement. Used the $t$-distribution with Satterthwaite $df$.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>cses</code></td>
<td>1.4710</td>
<td>3.0936</td>
</tr>
<tr>
<td><code>female</code></td>
<td>-1.5537</td>
<td>0.8734</td>
</tr>
<tr>
<td><code>minority</code></td>
<td>-5.9614</td>
<td>-2.4521</td>
</tr>
<tr>
<td><code>meanses</code></td>
<td>2.9316</td>
<td>5.5098</td>
</tr>
<tr>
<td><code>size</code></td>
<td>0.000317</td>
<td>0.001933</td>
</tr>
<tr>
<td><code>sector</code></td>
<td>0.6608</td>
<td>2.8112</td>
</tr>
<tr>
<td><code>cses*size</code></td>
<td>-0.00049</td>
<td>0.000555</td>
</tr>
<tr>
<td><code>cses*sector</code></td>
<td>-1.6548</td>
<td>-0.3518</td>
</tr>
<tr>
<td><code>female*meanses</code></td>
<td>-1.2785</td>
<td>1.2144</td>
</tr>
<tr>
<td><code>female*size</code></td>
<td>-0.00149</td>
<td>0.000080</td>
</tr>
<tr>
<td><code>female*sector</code></td>
<td>-1.4044</td>
<td>0.8032</td>
</tr>
<tr>
<td><code>minority*meanses</code></td>
<td>-2.1683</td>
<td>0.6098</td>
</tr>
<tr>
<td><code>minority*size</code></td>
<td>-0.00084</td>
<td>0.001208</td>
</tr>
<tr>
<td><code>minority*sector</code></td>
<td>0.7199</td>
<td>3.5179</td>
</tr>
</tbody>
</table>
For R Users

- Use the `lmerTest` package. Include "library(lmerTest)" and "require(lmerTest)". The `lmerTest` gives Satterwaite degrees of freedom and p-values for testing $\gamma_{k\ell} = 0$.

- There is a package that gives Kenward-Rogers.

- Alternatively you can compute confidence intervals using bootstrap, which completely avoids deciding on degrees of freedom. However, this take a very long time for complex models. I illustrate it using a simpler one:

```r
model1 <- lmer(match ~ 1 + cses + female + minority + meanses + sdsize + sector + (1 | id), data=hsb, REML=FALSE)
confint(model1, method='boot', nsim=1000, level=0.99)
```
### Results for bootstrap

<table>
<thead>
<tr>
<th>Effect</th>
<th>Lower 0.5%</th>
<th>Upper 99.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sqrt{\tau_0^2}$</td>
<td>0.96418282</td>
<td>1.5300571</td>
</tr>
<tr>
<td>.sig01</td>
<td>5.86039990</td>
<td>6.1315637</td>
</tr>
<tr>
<td>$\sqrt{\sigma^2}$</td>
<td>12.57027560</td>
<td>13.6542924</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>1.62953873</td>
<td>2.1609383</td>
</tr>
<tr>
<td>cses</td>
<td>−1.66326556</td>
<td>−0.8230537</td>
</tr>
<tr>
<td>female</td>
<td>−3.43482201</td>
<td>−2.3788735</td>
</tr>
<tr>
<td>minority</td>
<td>3.07801853</td>
<td>4.7141526</td>
</tr>
<tr>
<td>meanses</td>
<td>0.04007447</td>
<td>0.7368100</td>
</tr>
<tr>
<td>sdsizs</td>
<td>1.35047695</td>
<td>2.8932340</td>
</tr>
<tr>
<td>sector</td>
<td>1.35047695</td>
<td>2.8932340</td>
</tr>
</tbody>
</table>
Robust estimation: Why?

- When sample sizes are small, the Wald and $F$-tests can lead to different results. (HSB example: Large sample so differences were minor).

- If the random part of the model is wrong (i.e., non-normal data), then Wald and $F$-tests are not valid.

- Recall that the Wald and $F$ ($& t$) tests require:
  - The model for the mean of $Y_j$ is correctly specified, (i.e., $X_j \Gamma$) so that $E(\hat{\Gamma}) = \Gamma$ (i.e, unbiased).
  - The marginal covariance matrix is correctly specified, (i.e., $V_j = Z_j T Z_j' + \sigma^2 I$) so that the covariance matrix of the data equals the predicted covariance matrix.
Robust Estimation: What?

- **Problem**: If the random part of the model is wrong, then the results of Wald and $F$-tests are not valid.

- **Possible Solutions**:
  - Jackknife is OK but not as efficient as . . .
  - Bootstrap is computationally intense (e.g., R took a long time).
  - “Sandwich estimator” of the covariance matrix (Huber, 1967; White, 1982; see also Liang & Zeger, 1986).
Sandwich Estimator

- Uses the covariance matrices of the total residuals (i.e., total residuals $= y_j - X_j \hat{\Gamma}$) rather than the covariance matrices of the data (i.e., the $Y_j$’s).

- The sandwich estimator is also called the “robust” or the “empirical” variance estimator.

- It is consistent so long as the mean is correctly specified.
More Specifically What It Is

Recall (page 9),

\[
\text{cov}(\hat{\Gamma}) = \left[ (X'V^{-1}X)^{-1} \right] X'V^{-1}\Sigma_Y V^{-1}X \left[ (X'V^{-1}X)^{-1} \right] = M'\Sigma_Y M
\]

- Replace \( \Sigma_Y \) with

\[(y - X\hat{\Gamma})(y - X\hat{\Gamma})',\]

which is a block diagonal matrix with \((y_j - X_j\hat{\Gamma})(y_j - X_j\hat{\Gamma})'\) on the diagonal.

- The Sandwich estimator is consistent even if data are not normal (i.e., when model based one is inaccurate and inconsistent).

- If assumptions are met, Model Based estimator is more efficient.
Implications for practice

**Extreme point of view:**
If you’re only interested in the average (mean structure) in your data, then

- Ignore the within group dependency and use ordinary least squares to estimate the regression model.
- For inference, use the sandwich estimator, which corrects for within group dependency.

**Appropriate covariance model helps:**
- Interpretation and explanation of the random variation in the data.
- Improved efficiency (good for statistical inference).
- In longitudinal data analysis with missing data, the sandwich estimator is only appropriate if observations are missing at random.
### Simulation

<table>
<thead>
<tr>
<th>Parm</th>
<th>est</th>
<th>se</th>
<th>est</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{00}$</td>
<td>0.8201</td>
<td>0.8201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{10}$</td>
<td>-0.1018</td>
<td>-0.1018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{11}$</td>
<td>0.8581</td>
<td>0.8581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{20}$</td>
<td>-0.1563</td>
<td>-0.1563</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{21}$</td>
<td>0.1157</td>
<td>0.1157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{22}$</td>
<td>1.1409</td>
<td>1.1409</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>3.9115</td>
<td>3.9115</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>5.1656</td>
<td>0.1141</td>
<td>5.1656</td>
<td>0.1141</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>2.0524</td>
<td>0.0960</td>
<td>2.0524</td>
<td>0.0960</td>
</tr>
<tr>
<td>$\gamma_{20}$</td>
<td>3.0058</td>
<td>0.1086</td>
<td>3.0058</td>
<td>0.1086</td>
</tr>
</tbody>
</table>
## Simulation

<table>
<thead>
<tr>
<th>Effect</th>
<th>Correct Model</th>
<th></th>
<th></th>
<th>Wrong Model</th>
<th></th>
<th></th>
<th>sandwich</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>se</td>
<td>est</td>
<td>se</td>
<td>est</td>
<td>se</td>
<td></td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{00}$</td>
<td>0.8201</td>
<td>0.5352</td>
<td></td>
<td></td>
<td>0.5352</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{10}$</td>
<td>-0.1018</td>
<td>-0.0477</td>
<td></td>
<td></td>
<td>-0.0477</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{11}$</td>
<td>0.8581</td>
<td>0.8055</td>
<td></td>
<td></td>
<td>0.8055</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{20}$</td>
<td>-0.1563</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{21}$</td>
<td>0.1157</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{22}$</td>
<td>1.1409</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>3.9115</td>
<td>21.0821</td>
<td></td>
<td></td>
<td>21.0821</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>5.1656</td>
<td>0.1141</td>
<td></td>
<td>5.0903</td>
<td>0.1686</td>
<td></td>
<td>5.0903</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>2.0524</td>
<td>0.0960</td>
<td></td>
<td>2.0112</td>
<td>0.1035</td>
<td></td>
<td>2.0112</td>
</tr>
<tr>
<td>$\gamma_{20}$</td>
<td>3.0058</td>
<td>0.1086</td>
<td></td>
<td>3.1108</td>
<td>0.0387</td>
<td></td>
<td>3.1108</td>
</tr>
</tbody>
</table>
## Simulation

<table>
<thead>
<tr>
<th>Effect</th>
<th>Correct Model</th>
<th>Even Worse Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>est</td>
<td>se</td>
</tr>
<tr>
<td><strong>Random effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_{00}$</td>
<td>0.8201</td>
<td>1.3180</td>
</tr>
<tr>
<td>$\tau_{10}$</td>
<td>-0.1018</td>
<td></td>
</tr>
<tr>
<td>$\tau_{11}$</td>
<td>0.8581</td>
<td></td>
</tr>
<tr>
<td>$\tau_{20}$</td>
<td>-0.1563</td>
<td></td>
</tr>
<tr>
<td>$\tau_{21}$</td>
<td>0.1157</td>
<td></td>
</tr>
<tr>
<td>$\tau_{22}$</td>
<td>1.1409</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>3.9115</td>
<td>28.3097</td>
</tr>
<tr>
<td><strong>Fixed effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{00}$</td>
<td>5.1656</td>
<td>0.1141</td>
</tr>
<tr>
<td>$\gamma_{10}$</td>
<td>2.0524</td>
<td>0.0960</td>
</tr>
<tr>
<td>$\gamma_{20}$</td>
<td>3.0058</td>
<td>0.1086</td>
</tr>
</tbody>
</table>


**Eg. of Robust/Empirical Estimation**

Specify the “empirical” option in the PROC MIXED statement.

```latex
PROC MIXED data=hsbcent covtest method=ML empirical;
CLASS id;
MODEL mathach = cSES female minority meanSES size sector cSES*size cSES*sector female*meanSES female*size female*sector minority*meanSES minority*size minority*sector /solution chisq cl alpha=.01;
RANDOM intercept female minority cSES / subject=id type=un;
```
Model Information

Data Set: WORK.HSBCENT
Dependent Variable: mathach
Covariance Structure: Unstructured
Subject Effect: id
Estimation Method: ML
Residual Variance Method: Profile
Fixed Effects SE Method: Empirical
Degrees of Freedom Method: Containment
### Solution for Fixed Effects

| Effect       | Estimate | Error  | DF  | Value | Pr>|t| |
|--------------|----------|--------|-----|-------|-----|
| Intercept    | 12.0260  | 0.4269 | 155 | 28.17 | < .0001 |
| cses         | 2.2823   | 0.3176 | 157 | 7.19  | < .0001 |
| female       | -0.3402  | 0.4137 | 121 | -0.82 | 0.4126 |
| minority     | -4.2068  | 0.6439 | 133 | -6.53 | < .0001 |
| meanses      | 4.2207   | 0.4961 | 6604| 8.51  | < .0001 |
| size         | 0.001125 | 0.000296| 6604| 3.80  | 0.0001 |
| sector       | 1.7360   | 0.3978 | 6604| 4.36  | < .0001 |
| cses*size    | 0.000032 | 0.000222| 6604| 0.15  | 0.8837 |
| cses*sector  | -1.0033  | 0.2503 | 6604| -4.01 | < .0001 |
| female*meanses| -0.03207| 0.4235 | 6604| -0.08 | 0.9396 |
| female*size  | -0.00070 | 0.000256| 6604| -2.75 | 0.0059 |
| female*sector| -0.3006  | 0.4150 | 6604| -0.72 | 0.4689 |
| minority*meanses| -0.7793| 0.4933 | 6604| -1.58 | 0.1142 |
| minority*size | 0.000183 | 0.000386| 6604| 0.47  | 0.6349 |
| minority*sector| 2.1189  | 0.5398 | 6604| 3.93  | < .0001 |
## Contrasts with Robust Estimations

<table>
<thead>
<tr>
<th>Label</th>
<th>Num DF</th>
<th>Den DF</th>
<th>Chi-Square</th>
<th>$F$</th>
<th>Pr $&gt;\text{ChiSq}$</th>
<th>Pr $&gt; F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-level</td>
<td>5</td>
<td>6604</td>
<td>3.32</td>
<td>0.66</td>
<td>.6503</td>
<td>.6503</td>
</tr>
</tbody>
</table>
## Comparison: Model versus Robust

<table>
<thead>
<tr>
<th>Effect</th>
<th>Robust Estimation</th>
<th>Model-Based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Intercept</td>
<td>12.0260</td>
<td>0.4269</td>
</tr>
<tr>
<td>cses</td>
<td>2.2823</td>
<td>0.3176</td>
</tr>
<tr>
<td>female</td>
<td>−0.3402</td>
<td>0.4137</td>
</tr>
<tr>
<td>minority</td>
<td>−4.2068</td>
<td>0.6439</td>
</tr>
<tr>
<td>meanses</td>
<td>4.2207</td>
<td>0.4961</td>
</tr>
<tr>
<td>size</td>
<td>0.001125</td>
<td>0.000296</td>
</tr>
<tr>
<td>sector</td>
<td>1.7360</td>
<td>0.3978</td>
</tr>
<tr>
<td>cses*size</td>
<td>0.000032</td>
<td>0.000222</td>
</tr>
<tr>
<td>cses*sector</td>
<td>−1.0033</td>
<td>0.2503</td>
</tr>
<tr>
<td>female*meanses</td>
<td>−0.03207</td>
<td>0.4235</td>
</tr>
<tr>
<td>female*size</td>
<td>−0.00070</td>
<td>0.000256</td>
</tr>
<tr>
<td>female*sector</td>
<td>−0.3006</td>
<td>0.4150</td>
</tr>
<tr>
<td>minority*meanses</td>
<td>−0.7793</td>
<td>0.4933</td>
</tr>
<tr>
<td>minority*size</td>
<td>0.000183</td>
<td>0.000386</td>
</tr>
<tr>
<td>minority*sector</td>
<td>2.1189</td>
<td>0.5398</td>
</tr>
</tbody>
</table>
For R Users: Empirical SEs

- `lmer` does not compute these and as far as I as could tell there is no packager that computes them.

- I wrote a function called `robust` that will compute them:
  - Fit a model, `modelA`
    ```r
    modelA <- lmer(mathach ~ 1 + cSES + female + minority + meanses + ssize + sector + cSES*ssize + cSES*sector + female*meanses + female*ssize + female*sector + minority*meanses + minority*sector + minority*ssize + (1 + cSES + female | id), data = hsb, REML = FALSE)
    ```
  - Run the function `robust` in the console window
  - To get robust standard errors, type
    ```r
    robust(modelA, hsb$mathach, hsb$id, dftype)
    ```
    where `dftype` equals either residual or between/within

- Output includes: effect, model bases se's, and robust se's.
Comparison: Model versus Robust

Notes:

- Estimates of fixed effect are exactly the same.
- Estimates of SE’s differ a little.
- If you miss-specify the mean structure, the SE’s differ more.
- If you miss-specify the random structure, the SE’s differ more.
- We’ll stick to model-based because we’re interested in random effects; however, it can be a good thing to use robust when model building.
The Classic: Likelihood Ratio Tests

- The classical statistical test for comparing nested models.
- Suppose that we have two models that have the same fixed and random effects, except one model has $\gamma_{kl} = 0$.
- The Full Model is the one with all the parameters.
- The Reduced model is the one with $\gamma_{kl} = 0$.
- Likelihood ratio test for

$$H_0 : \gamma_{kl} = 0 \quad \text{versus} \quad H_a : \gamma_{kl} \neq 0$$
**Likelihood Ratio Test Statistic**

is defined as

\[-2 \ln \lambda_N = -2 \ln \left[ \frac{L_{ML}(\hat{\Gamma}_o, \hat{T}, \hat{\sigma}^2)}{L_{ML}(\hat{\Gamma}, \hat{T}, \hat{\sigma}^2)} \right] = -2(\ln[L_{ML}(\hat{\Gamma}_o, \hat{T}, \hat{\sigma}^2)] - \ln[L_{ML}(\hat{\Gamma}, \hat{T}, \hat{\sigma}^2)])\]

where

- $L_{ML}(\hat{\Gamma}_o, \hat{T}, \hat{\sigma}^2) =$ the value of the likelihood function under the nested model.

- $L_{ML}(\hat{\Gamma}, \hat{T}, \hat{\sigma}^2) =$ the value of the likelihood function under the full model.
Likelihood Ratio Test

If $H_o$ is true (as well as all other assumptions),

Then $LR$ is asymptotically distributed as a $\chi^2$ random variable with degrees of freedom equal to the difference between the number of $\gamma$'s in the two models.

The likelihood ratio test for fixed effects is only valid for ML estimation.
The LR test is not valid under REML.

- Recall that in REML
  1. Remove the mean structure from the data & then estimate the covariance matrix for the random effects.
  2. Given $\hat{T}$ & $\hat{\sigma}^2$, use standard estimation techniques to estimate the mean structure (i.e., the $\gamma$'s).

- Under REML, two models with different mean structures have likelihood functions based on different observations so the likelihoods are not comparable.
Example of Likelihood Ratio Test

LR test on the set of cross-level interactions where the statistical hypothesis is

$$H_0 : \mathbf{L}\Gamma = \begin{pmatrix} \gamma_{12} \\ \gamma_{21} \\ \gamma_{23} \\ \gamma_{32} \\ \gamma_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

versus

$$H_a : \mathbf{L}\Gamma \neq 0$$
Example of Likelihood Ratio Test (continued)

For the likelihood ratio test, we compute the model with and without these effects and record $-2 \ln(\text{likelihood})$:

<table>
<thead>
<tr>
<th>Model</th>
<th>ML</th>
<th>REML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced or null</td>
<td>46,223.5645</td>
<td>46,263.2394</td>
</tr>
<tr>
<td>Full</td>
<td>46,220.8436</td>
<td>46,288.5541</td>
</tr>
</tbody>
</table>

$-2 \ln \lambda_N = 2.7209$  \hspace{1cm} $-25.3147$

$df = 5$

$p$-value = .74
## Summary: Tests for Fixed Effects

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Value</th>
<th>Distribution</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model Based</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>2.97</td>
<td>$\chi^2_5$</td>
<td>.70</td>
</tr>
<tr>
<td>$F$</td>
<td>.59</td>
<td>$F_{5,6604}$</td>
<td>.70</td>
</tr>
<tr>
<td>$-2\ln \lambda_N$</td>
<td>2.72</td>
<td>$\chi^2_5$</td>
<td>.74</td>
</tr>
<tr>
<td><strong>Robust Estimation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald</td>
<td>3.32</td>
<td>$\chi^2_5$</td>
<td>.65</td>
</tr>
<tr>
<td>$F$</td>
<td>.66</td>
<td>$F_{5,6604}$</td>
<td>.65</td>
</tr>
</tbody>
</table>
Before Tests for Variance Components

Simplify by dropping the 5 cross-level interactions.

**Level 1**

\[
(math)_{ij} = \beta_{0j} + \beta_{1j}(cSES)_{ij} + \beta_{2j}(female)_{ij} + \beta_{3j}(minority)_{ij} + R_{ij}
\]

where \( R_{ij} \sim \mathcal{N}(0, \sigma^2) \) i.i.d.

**Level 2**

\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(sector)_j + \gamma_{02}(size)_j + \gamma_{03}(SES)_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(sector)_j + U_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}(size)_j + U_{2j} \\
\beta_{3j} &= \gamma_{30} + \gamma_{31}(sector)_j + U_{3j}
\end{align*}
\]
Linear Mixed Model

\[
(math)_{ij} = \left[ \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\text{SES})_j \right] \\
+ \left[ \gamma_{10} + \gamma_{11}(\text{sector})_j \right](\text{cSES})_{ij} \\
+ \left[ \gamma_{20} + \gamma_{21}(\text{size})_j \right](\text{female})_{ij} \\
+ \left[ \gamma_{30} + \gamma_{31}(\text{sector})_j \right](\text{minority})_{ij} \\
+ U_{0j} + U_{1j}(\text{cSES})_{ij} + U_{2j}(\text{female})_{ij} + U_{3j}(\text{minority})_{ij} \\
+ R_{ij}
\]

\[
= \gamma_{00} + \gamma_{10}(\text{cSES})_{ij} + \gamma_{20}(\text{female})_{ij} + \gamma_{30}(\text{minority})_{ij} \\
+ \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\text{SES})_j \\
+ \gamma_{11}(\text{sector})_j(\text{cSES})_{ij} + \gamma_{21}(\text{size})_j(\text{female})_{ij} \\
+ \gamma_{31}(\text{sector})_j(\text{minority})_{ij} + U_{0j} \\
+ U_{1j}(\text{cSES})_{ij} + U_{2j}(\text{female})_{ij} + U_{3j}(\text{minority})_{ij} + R_{ij}
\]
# Simpler Model: Model Information

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
<td>WORK.HSBCENT</td>
</tr>
<tr>
<td>Dependent Variable</td>
<td>mathach</td>
</tr>
<tr>
<td>Covariance Structure</td>
<td>Unstructured</td>
</tr>
<tr>
<td>Subject Effect</td>
<td>id</td>
</tr>
<tr>
<td>Estimation Method</td>
<td>ML</td>
</tr>
<tr>
<td>Residual Variance Method</td>
<td>Profile</td>
</tr>
<tr>
<td>Fixed Effects SE Method</td>
<td>Model-Based</td>
</tr>
<tr>
<td>Degrees of Freedom Method</td>
<td>Satterthwaite</td>
</tr>
</tbody>
</table>

Convergence criteria met.
## Solution for Fixed Effects

| Effect            | Estimate | Error   | DF | Value | Pr>|t| |
|-------------------|----------|---------|----|-------|----|-----|
| Intercept         | 12.0837  | 0.4123  | 175| 29.31 | < .0001 |
| cses              | 2.3240   | 0.1518  | 151| 15.31 | < .0001 |
| female            | −0.5404  | 0.3588  | 138| −1.51 | 0.1343 |
| minority          | −3.7925  | 0.3135  | 174| −12.10| < .0001 |
| meanses           | 3.9813   | 0.3298  | 155| 12.07 | < .0001 |
| size              | 0.001132 | 0.000288| 174| 3.92  | 0.0001 |
| sector            | 1.6179   | 0.2965  | 128| 5.46  | < .0001 |
| cses*sector       | −1.0115  | 0.2263  | 153| −4.47 | < .0001 |
| female*size       | −0.00062 | 0.000279| 144| −2.24 | 0.0269 |
| minority*sector   | 1.7647   | 0.4321  | 126| 4.08  | < .0001 |
Estimated Structural Model...

\[
\hat{y}_{ij} = [12.084 + 1.618(\text{sector})_j \\
+ 0.001(\text{size})_j + 3.98(\text{SES})_j] \\
+ [2.324 - 1.012(\text{sector})_j](\text{cSES})_{ij} \\
+ [-0.540 - 0.001(\text{size})_j](\text{female})_{ij} \\
+ [-3.793 + 1.765(\text{sector})_j](\text{minority})_{ij}
\]

...for now...
Inference for Variance Components

Need adequate covariance matrix for the random effects (i.e., $T$) because

- Useful for interpreting random variation in the data.
- Essential for model-based inferences.
  - Over-parameterization of covariance structure $\rightarrow$ inefficient (and possibly poor) estimated standard errors for the fixed effects.
  - Under-parameterization of covariance structure $\rightarrow$ invalid inferences for the fixed effects.
Inference for Variance Components

- Approximate Wald tests (z tests).
- Likelihood ratio tests.
- Testing the number of random effects.
Approximate Wald Tests

- For both ML and REML.

- For the **marginal model**, variance components are asymptotic normal with the covariance matrix given by \((-H)^{-1}\), where \(H\) is the Hessian.

- Wald tests (& confidence statements) for:
  1. **Variances**, i.e.,
     \[ H_o : \tau_k^2 = 0 \quad \text{versus} \quad H_a : \tau_k^2 \neq 0 \]
  2. **Covariances**, e.g.,
     \[ H_o : \tau_{kl} = 0 \quad \text{versus} \quad H_a : \tau_{kl} \neq 0 \quad \text{for} \ k \neq l \]
Approximate Wald Tests: Variances

For example: \( H_0 : \tau_k^2 = 0 \) versus \( H_a : \tau_k^2 \neq 0 \)

- The closer \( \tau_k^2 \) is to 0, the larger the sample needed for approximate normality to hold.

- Whether the model is marginal or hierarchical now becomes very important —

For a hierarchical linear model, the variances of random effects cannot be negative. If \( \tau_k^2 = 0 \), then the normal approximation completely fails because a variance \( \tau_k^2 \) cannot be non-negative.
Variance: Wald Test Statistic

\[ z = \frac{\hat{\tau}^2_k}{S.E.} \]

Example: HSB data and SAS/MIXED commands:

```sas
PROC MIXED data=hsbcent noclprint covtest method=ML;
  CLASS id;
  MODEL mathach = cSES female minority meanSES size cSES*sector female*size minority*sector / solution chisq ddfm=satterth;
  RANDOM intercept female minority cSES / subject=id type=un;
```
## Variance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Z</th>
<th>Pr &gt; Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>UN(1,1)</td>
<td>id</td>
<td>2.2408</td>
<td>0.4991</td>
<td>4.49</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>UN(2,2)</td>
<td>id</td>
<td>0.6791</td>
<td>0.5117</td>
<td>1.33</td>
<td>0.0922</td>
</tr>
<tr>
<td>UN(3,3)</td>
<td>id</td>
<td>0.9088</td>
<td>0.6936</td>
<td>1.31</td>
<td>0.0951</td>
</tr>
<tr>
<td>UN(4,4)</td>
<td>id</td>
<td>0.1412</td>
<td>0.2118</td>
<td>0.67</td>
<td>0.2525</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>35.3169</td>
<td>0.6106</td>
<td>57.84</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>
Covariances

- For example,
  \[ H_0 : \tau_{kl} = 0 \quad \text{for} \quad k \neq l \quad \text{versus} \quad H_a : \tau_{kl} \neq 0 \]

- The distinction between marginal model and HLM (random effects model) is less crucial.

- For a valid test for the covariances, still need to assume that all \( \tau_k^2 \)'s are greater than 0.

- SAS/MIXED results for covariances (and variances) . . .
### Covariances Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Subject</th>
<th>Estimate</th>
<th>Error</th>
<th>Value</th>
<th>Pr Z</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>UN(1,1)</em></td>
<td>id</td>
<td>2.2408*</td>
<td>0.4991*</td>
<td>4.49*</td>
<td>&lt; .0001*</td>
</tr>
<tr>
<td>UN(2,1)</td>
<td>id</td>
<td>−0.9391</td>
<td>0.4358</td>
<td>−2.15</td>
<td>0.0312</td>
</tr>
<tr>
<td><em>UN(2,2)</em></td>
<td>id</td>
<td>0.6791*</td>
<td>0.5117*</td>
<td>1.33*</td>
<td>0.0922*</td>
</tr>
<tr>
<td>UN(3,1)</td>
<td>id</td>
<td>−0.1530</td>
<td>0.5090</td>
<td>−0.30</td>
<td>0.7638</td>
</tr>
<tr>
<td>UN(3,2)</td>
<td>id</td>
<td>0.2106</td>
<td>0.4712</td>
<td>0.45</td>
<td>0.6548</td>
</tr>
<tr>
<td><em>UN(3,3)</em></td>
<td>id</td>
<td>0.9088*</td>
<td>0.6936*</td>
<td>1.31*</td>
<td>0.0951*</td>
</tr>
<tr>
<td>UN(4,1)</td>
<td>id</td>
<td>0.1467</td>
<td>0.2576</td>
<td>0.57</td>
<td>0.5690</td>
</tr>
<tr>
<td>UN(4,2)</td>
<td>id</td>
<td>−0.1163</td>
<td>0.2395</td>
<td>−0.49</td>
<td>0.6274</td>
</tr>
<tr>
<td>UN(4,3)</td>
<td>id</td>
<td>−0.2376</td>
<td>0.2965</td>
<td>−0.80</td>
<td>0.4230</td>
</tr>
<tr>
<td><em>UN(4,4)</em></td>
<td>id</td>
<td>0.1412*</td>
<td>0.2118*</td>
<td>0.67*</td>
<td>0.2525*</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>35.3169</td>
<td>0.6106</td>
<td>57.84</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

“*” indicates statistics for a variance.
Likelihood Ratio Test for Variances?

The Likelihood ratio test statistic for variance components is

\[-2 \ln \lambda_N = -2 \ln \left( \frac{L_{ML}(\hat{\Gamma}, \hat{T}_o, \hat{\sigma}^2)}{L_{ML}(\hat{\Gamma}, \hat{T}, \hat{\sigma}^2)} \right) = -2(\ln[L_{ML}(\hat{\Gamma}, \hat{T}_o, \hat{\sigma}^2)] - \ln[L_{ML}(\hat{\Gamma}, \hat{T})]),\]

where

- \( L_{ML}(\hat{\Gamma}, \hat{T}_o, \hat{\sigma}^2) = \) the value of the likelihood function under the nest model.

- \( L_{ML}(\hat{\Gamma}, \hat{T}, \hat{\sigma}^2) = \) the value of the likelihood function under the full model.
Likelihood Ratio Test Variances?

- You can use REML or ML (unlike the fixed effects case).
- The test statistic has an approximate $\chi^2$ distribution with degrees of freedom equal to the difference in the number of parameters between the nested and full models.
- One of the required conditions ("regularity conditions") that gives the distribution for the test statistic is that the parameter estimates are not on the boundary of the parameter space. Therefore, . . .
- For the HLM, the likelihood ratio test is not valid if $\tau_k^2 = 0$.
- For the marginal model, the likelihood ratio test is fine.

Since the Wald and Likelihood ratio tests are not valid when $\tau_k^2 = 0$, we use an alternative to approach to evaluate $H_0: \tau_k^2 = 0$. 
Testing the Number of Random Effects

- Goal is to test whether we need (some) of the random effects. e.g., Whether we need a random slope for cSES in the HSB example:

  \[ H_0 : \tau_{30} = \tau_{31} = \tau_{32} = \tau_{33}^2 = 0. \]

- When a \( \tau_k^2 = 0 \) is on boundary of the parameter space, so we can’t use the Wald or the likelihood ratio test and compare the test statistic to a Chi-square distribution.
Testing the Number of Random Effects

The test that we can do is based on


Testing the Number of Random Effects

The test statistic is the likelihood ratio test statistic, but sampling distribution of the test statistic is a mixture of two $\chi^2$ distributions.

Before presenting general rules, we’ll consider 4 cases:

- No random effects versus one random effect (i.e., random intercept).
- One versus Two Random effects.
- $q$ versus $q + 1$ random effects.
- $q$ versus $q + k$ random effects.
Case 1: One versus Two Random effects.

This is essentially testing for a random intercept:

\[ H_0 : \tau_0^2 = 0 \quad \text{versus} \quad \tau_0^2 > 0 \]

If \( H_0 \) is true, then the distribution of

\[ -2 \ln \lambda_N = -2(\ln[L_{ML}(\hat{\Gamma}, \hat{T}_o, \hat{\sigma}^2)] - \ln[L_{ML}(\hat{\Gamma}, \hat{T}, \hat{\sigma}^2)]) \]

is a mixture of \( \chi^2_1 \) and \( \chi^2_0 \) distributions where we give equal weights to each (i.e., 1/2).
Mixture of $\chi^2_0$ & $\chi^2_1$ with Equal Weights

Mixture of $\chi^2_1$ and $\chi^2_0$
Example of Case 1: HSB (using ML)

**Null model**: no random effects

**Full model**: random intercept.

<table>
<thead>
<tr>
<th>Model of τ’s</th>
<th>Deviance ( -2 \ln(\lambda_N) )</th>
<th>Test statistic</th>
<th>( \chi^2_0 )</th>
<th>( \chi^2_1 )</th>
<th>p-value from mixture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>0</td>
<td>46,372.3</td>
<td>137.6</td>
<td>0</td>
<td>.89E − 31</td>
</tr>
<tr>
<td>Full</td>
<td>1</td>
<td>46,234.7</td>
<td>—</td>
<td>.45E − 31</td>
<td>.45E − 31</td>
</tr>
</tbody>
</table>

The mixture p-value = \(.5(\ .89E − 31) = .45E − 31\). 

C.J. Anderson (Illinois)  
Statistical Inference: The Marginal Model  
Fall 2017  
86.1/ 130
Mixture of $\chi^2_0$ & $\chi^2_1$ with Equal Weights
Case 2: One vs Two Random Effects

\[ H_0 : \mathbf{T} = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & 0 \end{pmatrix} \text{ versus } H_a : \mathbf{T} = \begin{pmatrix} \tau_0^2 & \tau_{10} \\ \tau_{10} & \tau_1^2 \end{pmatrix} \]

In other words, we’re testing

\[ H_0 : \tau_{10} = \tau_1^2 = 0 \text{ versus } H_a : \text{not } \tau_{10} = \tau_1^2 = 0 \]

Assuming

- \( \tau_0^2 > 0 \) in \( H_0 \)

- In \( H_a \), \( \mathbf{T} \) is a “proper” covariance matrix
  (i.e., \( \tau_{10} \leq \tau_1 \tau_0 \), and \( \tau_k^2 > 0 \)).

To get the correct \( p \)-value, we take a mixture of \( \chi_1^2 \) and \( \chi_2^2 \) distributions.
To get the correct $p$-value we take a mixture of $\chi^2_1$ and $\chi^2_2$ distributions.
Case 2: HSB Example

**Null model:** Random intercept only

**Full model:** Random intercept and random slope for “female”

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of $\tau$’s</th>
<th>$-2 \ln(\lambda_N)$</th>
<th>Test statistic</th>
<th>$p$-value from $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>1</td>
<td>46,234.68</td>
<td>5.40</td>
<td>.020</td>
</tr>
<tr>
<td>Full</td>
<td>3</td>
<td>46,229.28</td>
<td>—</td>
<td>.067</td>
</tr>
</tbody>
</table>

Mixture $p$-value = $.5(.020) + .5(.067) = .04.$

Note Wald $p = .09.$
Mixture of $\chi^2_1$ & $\chi^2_2$ with Equal Weights
Case 3: \( q \) vs \( q + 1 \) Random Effects

The hypotheses are

\[
H_o : \quad T = \begin{pmatrix}
\tau_0^2 & \tau_{10} & \cdots & \tau_{q0} & 0 \\
\tau_{10} & \tau_1^2 & \cdots & \tau_{q1} & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\tau_{q0} & \tau_{q1} & \cdots & \tau_{qq} & 0 \\
0 & 0 & \cdots & 0 & 0
\end{pmatrix}
\]

and

\[
H_a : \quad T = \begin{pmatrix}
\tau_0^2 & \tau_{10} & \cdots & \tau_{q0} & \tau_{(q+1)0} \\
\tau_{10} & \tau_1^2 & \cdots & \tau_{q1} & \tau_{(q+1)1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\tau_{q0} & \tau_{q1} & \cdots & \tau_{qq} & \tau_{(q+1)q} \\
\tau_{(q+1)0} & \tau_{(q+1)1} & \cdots & \tau_{(q+1)q} & \tau_{(q+1)}^2
\end{pmatrix}
\]
Case 3 (continued)

Assuming that

- In $H_0$, the $(q \times q)$ matrix of $\tau$’s is a “proper” covariance matrix.

- In $H_a$, the $((q + 1) \times (q + 1))$ matrix is a “proper” covariance matrix.

Then the asymptotic sampling distribution of $-2 \ln(\lambda_N)$ is a mixture of $\chi^2_q$ and $\chi^2_{q+1}$. 
Case 3: HSB Example

- **Null model**: Random intercept and random slopes for “female” and “minority”.
- **Full model**: Random intercept and random slopes for “female”, “minority” and “cSES”.

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of $\tau$’s</th>
<th>$-2 \ln(\lambda_N)$</th>
<th>Test statistic</th>
<th>$\chi_3^2$</th>
<th>$\chi_4^2$</th>
<th>Mixture $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>6</td>
<td>46,225.12</td>
<td>1.55</td>
<td>.67</td>
<td>.82</td>
<td>.74</td>
</tr>
<tr>
<td>Full</td>
<td>10</td>
<td>46,223.56</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Mixture $p$-value = $0.5(0.67) + 0.5(0.82) = 0.74.$
Mixture of $\chi^2_3 \& \chi^2_4$ with Equal Weights
Case 4: $q$ vs $q + k$ Random Effects

- The sample distribution of $-2 \ln(\lambda_N)$ is a mixture of $\chi^2$ random variables and other random variables.

- Based on semi-current statistical knowledge, getting $p$-values for this case requires simulations to estimate the appropriate sampling distribution of the test statistic.
<table>
<thead>
<tr>
<th>Model Description</th>
<th>$-2 \ln(\lambda_N)$</th>
<th>$q$</th>
<th>Correct P-Value</th>
<th>Naive P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, Minority, cSES</td>
<td>46, 233.56</td>
<td>10</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>Female, Minority</td>
<td>46, 225.12</td>
<td>6</td>
<td>0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Female</td>
<td>46, 229.23</td>
<td>3</td>
<td>0.18</td>
<td>0.26</td>
</tr>
<tr>
<td>Minority</td>
<td>46, 231.95</td>
<td>3</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>cSES</td>
<td>46, 233.98</td>
<td>6</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Intercept Only</td>
<td>46, 234.68</td>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>No Random Effects</td>
<td>46, 372.30</td>
<td>0</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>
Summary of The General Procedure

To test $q$ versus $q + 1$ random effects:

$$H_0 : \tau_{q+1}^2 = \tau_{q,q+1} = \ldots = \tau_{0,q+1} = 0 \text{ vs } H_a : \text{not } H_0$$

- $T$ must be a proper covariance matrix (i.e., $\tau_k^2 > 0$ and $\tau_{kk'} \geq \tau_k \tau_{k'}$).
- Fit nested and full model.
- Compute likelihood ratio test statistic.
- Compare test statistic to $\chi^2_q$ and $\chi^2_{q-1}$.
- Average the $p$-values.
Summary Comments

The validity of statistical tests for number of random effects depend on

1. The likelihood function being maximized over a parameter space where \( \tau_{kl} \leq \tau_k \tau_l \) and \( \tau_k^2 \geq 0 \).

In linear algebra terms, \( \mathbf{T} \) is “positive semi-definite,” that is, it is a “proper” covariance matrix.

2. The estimating procedure converges.

Note: The first condition regarding the parameter space is software dependent —

In SAS/MIXED, the parameter space is bigger than necessary; that is, we can get \( \tau_{kl} > \tau_k \tau_l \). So need to check to make sure that \( \hat{\mathbf{T}} \) is a “proper” covariance matrix (i.e., no correlations \( \geq 1 \) or \( \leq -1 \)).
Summary Comments (continued)

on Tests for number of random effects

- The procedure described here differs from Snijders & Bosker (1999) (Section 6.2.1). Snijders & Bosker (1999) was based on Self & Liang (1987) and follows the results given by Stram & Lee (1994).

- When Stram & Lee (1994) wrote their paper, SAS/MIXED required $T$ to be “positive definite,” which is too restrictive for the mixture results. So they suggest corrections that consist of halving $p$-values, which is what Snijders & Bosker discuss in section 6.2.1.

- In the 2nd edition of Snijders & Bosker (2012) the correct procedure is given.
Global Measures of Fit

...and some statistics to use in model selection.

Those covered here

- Can be used to compare nested and/or non-nested models.
- Are not statistical tests of significance.
- Specifically,
  - Information criteria
  - $R^2$ type measures
Information Criteria

- They all start with the value of the likelihood function of a model and adjust it based on
  - Model complexity (i.e., number of parameters)
  - Sample size

When comparing models, all models should be estimated by MLE. If you are using REML, the only models that can be compared are those with the same fixed effects. Just as likelihoods for fixed effects are not comparable, ICs using these likelihoods are also not comparable.

- Four common ones (and ones that SAS/MIXED) computed.
Information Criteria (continued)

Let

- $\mathcal{L} =$ the maximum of the log of the likelihood function.

- $d =$ dimension of the model; that is, the number of estimated parameters. This includes all the $\gamma$’s, $\tau$’s and $\sigma^2$.

- $N$ or the “effective sample” size, which for us is the number of marco units.
## Four Information Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Smaller-is-better</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>$-2L + 2d$</td>
<td>Akaike (1974)</td>
</tr>
<tr>
<td>AAIC</td>
<td>$-2L + 2dn^<em>(n^</em> - d - 1)$</td>
<td>Hurvich &amp; Tsai (1989)</td>
</tr>
<tr>
<td>HQIC</td>
<td>$-2L + 2d \log \log N$</td>
<td>Hannan &amp; Quinn (1979)</td>
</tr>
<tr>
<td>BIC</td>
<td>$-2L + d \log N$</td>
<td>Schwarz (1978)</td>
</tr>
<tr>
<td>CAIC</td>
<td>$-2L + d \log(N + 1)$</td>
<td>Bozdogan (1987)</td>
</tr>
</tbody>
</table>
Notes Regarding Information Criteria

- Be sure that you know whether the one you’re using is a “larger” or “smaller–is–better.”
- SAS/MIXED gives you “smaller–is–better”
- To get (all of) these in the output, add “ic” as an option to “PROC MIXED”.
- Earlier versions of SAS/MIXED didn’t include $\sigma^2$ as a parameter.
- Information criteria are only “rules of thumb” and not statistical tests.
- The different criteria many not always agree.
**Example: HSB \( N = 160 \)**

Random intercept and different fixed effects.

<table>
<thead>
<tr>
<th>Model</th>
<th>(-2 \ln(\text{like}))</th>
<th>(q)</th>
<th>AIC</th>
<th>HQIC</th>
<th>BIC</th>
<th>CAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) null</td>
<td>47115.8</td>
<td>3</td>
<td>47121</td>
<td>47125</td>
<td>47131</td>
<td>47134</td>
</tr>
<tr>
<td>(b) cSES</td>
<td>46720.4</td>
<td>4</td>
<td>46728</td>
<td>46733</td>
<td>46740</td>
<td>46744</td>
</tr>
<tr>
<td>(c) (b) + minority</td>
<td>46513.1</td>
<td>5</td>
<td>46523</td>
<td>46529</td>
<td>46538</td>
<td>46543</td>
</tr>
<tr>
<td>(d) (c) + female</td>
<td>46456.9</td>
<td>6</td>
<td>46468</td>
<td>46476</td>
<td>46487</td>
<td>46493</td>
</tr>
</tbody>
</table>

Adding level 2 variables to model (d)

| (e) size, sector & meanSES | 46289.3 | 9 | 46307 | 46318 | 46335 | 46344 |
| (f) (e) + hinimity         | 46287.6 | 10| 46307 | 46320 | 46338 | 46348 |

Adding cross-level to model (e)

| (h) cSES*sector female*size minority*sector | 46234.7 | 12| 46258 | 46273 | 46295 | 46307 |
| (i) 5 more cross-level                | 46206.2 | 17| 46240 | 46261 | 46292 | 46309 |
Example: HSB $N = 160$ (continued)

Since the extra five cross-level interactions were not statistically significant (with a very complex random structure), we’ll go with model (h).

What about varying the random structure?

The following models all include the same fixed effects as model (h) and were fit via MLE.
### Example: HSB $N = 160$ (continued)

Different random structures:

<table>
<thead>
<tr>
<th>Model</th>
<th>params</th>
<th>AIC</th>
<th>HQIC</th>
<th>BIC</th>
<th>CAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept, female, cSES, &amp; minority</td>
<td>21</td>
<td>46265</td>
<td>46291</td>
<td>46330</td>
<td>46351</td>
</tr>
<tr>
<td>intercept, female &amp; minority</td>
<td>17</td>
<td>46259</td>
<td>46280</td>
<td>46311</td>
<td>46328</td>
</tr>
<tr>
<td>intercept, cSES &amp; minority</td>
<td>17</td>
<td>46264</td>
<td>46285</td>
<td>46316</td>
<td>46333</td>
</tr>
<tr>
<td>intercept, cSES &amp; female</td>
<td>17</td>
<td>46262</td>
<td>46283</td>
<td>46314</td>
<td>46331</td>
</tr>
<tr>
<td>intercept, female</td>
<td>14</td>
<td><strong>46257</strong></td>
<td>46274</td>
<td>46300</td>
<td>46314</td>
</tr>
<tr>
<td>intercept, minority</td>
<td>14</td>
<td>46260</td>
<td>46277</td>
<td>46303</td>
<td>46317</td>
</tr>
<tr>
<td>intercept, cSES</td>
<td>14</td>
<td>46262</td>
<td>46279</td>
<td>46305</td>
<td>46319</td>
</tr>
<tr>
<td>intercept</td>
<td>12</td>
<td>46258</td>
<td><strong>46273</strong></td>
<td><strong>46295</strong></td>
<td>46307</td>
</tr>
<tr>
<td>none</td>
<td>11</td>
<td>46371</td>
<td>46409</td>
<td>46481</td>
<td>46497</td>
</tr>
</tbody>
</table>
Summary:

**HSB (N = 160)**

1.55 (.74/.82)  
4.49 (.28/.34)  
7.04 (.10/.13)

- Female, minority, cSES  
  \(-2 \ln(\lambda_N) = 46, 225.12\)  
  \(AIC = 46, 259.1\)

4.17 (.18/.24)  
6.84 (.05/.08)

- Female  
  \(-2 \ln(\lambda_N) = 46, 229.23\)  
  \(AIC = 46, 257.3\)

5.40 (.04/.06)  
2.73 (.18/.26)  
.70 (.55/.70)

- Minority  
  \(-2 \ln(\lambda_N) = 46, 231.95\)  
  \(AIC = 46, 260.0\)

- cSES  
  \(-2 \ln(\lambda_N) = 46, 233.98\)  
  \(AIC = 46, 262.0\)

- Intercept only  
  \(-2 \ln(\lambda_N) = 46, 234.68\)  
  \(AIC = 46, 258.7\)

137.6 (< .0001)

- No random effects  
  \(-2 \ln(\lambda_N) = 46, 372.3\)  
  \(AIC = 46, 394.3\)

Model (h) w/ random intercept & random slopes for effects listed
Summary Comments: Info. Criteria

- Information criteria are only “rules of thumb” and not statistical tests.
- The different criteria may not always agree.
- Information criteria are different ways of making a subjective decision (i.e., selecting a good model) look objective.
- Model selection is a process of gathering evidence and doesn’t rest only any single statistic.
$R^2$ type measures

Extend concept from multiple regression $\longrightarrow R^2$.

Uses in the multi-level context:

- Indices of fit.
- Can be used for diagnostic purposes.
**R² type measures**

In multiple regression, there are different ways to derive $R^2$:

- The maximal squared correlation between the observed and predicted $Y$.
- The proportional reduction in unexplained (modeled) variance of $Y$ due to using predictor variables.
- The proportional reduction in prediction error variance.

They don’t all work with multilevel models.

---

**R² Measures in Multilevel Models:**

We need to consider micro and macro level residual variance.

So we need to propose measures for each level:

- $R^2_1$: level 1
- $R^2_2$: level 2
Proportional Reduction Unexplained Variance

For level 2: \( R_2^2 = \frac{\tau_0^{*2}}{\tau_0^2} \)

where

- \( \tau_0^{*2} \) is level 2 residual variance with predictor variables (micro and/or macro) in the model.
- \( \tau_0^2 \) is without predictor variables.

This value can be greater than one; that is, when \( \tau_0^{*2} > \tau_0^2 \), \( R_2^2 \Rightarrow 1 \). (see Snijders & Bosker for an example).

...a better approach... The \( R^2 \) measures for multilevel models are only appropriate (make sense) when data come from an observational study; that is, the predictor variables are random.

We’ll go over this for random intercept models:

- Level 1
- Level 2
Level 1: Proportional Reduction in Prediction Error

**Level 1:** We want a measure of the decrease in prediction error when predicting $Y_{ij}$, in particular, we want to predict $Y_{ij}$ for a randomly drawn individual $i$ from a randomly drawn group $j$.

Suppose that the (linear mixed) model in the population is

$$Y_{ij} = \sum_{k=0}^{p} \gamma_k X_{k,ij} + U_{0j} + R_{ij}$$

where $X_{0,ij} = 1$ for all individuals and groups.

The $X_k$ are random variables but we don’t know what they equal.
Level 1, Case 1

The prediction that will minimize the sum of squared errors is the expected value of $Y_{ij}$,

$$E(Y_{ij}) = E \left[ \sum_{k=0}^{p} \gamma_{k0}X_{k,ij} + U_{0j} + R_{ij} \right]$$

$$= \sum_{k=0}^{p} \gamma_{k0} E[X_{k,ij}] + E[U_{0j}] + E[R_{ij}]$$

$$E(Y_{ij}) = \sum_{k=0}^{p} \gamma_{k0} \mu_k$$

where

- The $\gamma$’s are fixed (considered to be known).
- The $X_k$’s are random variables with means $\mu_k$.
- The random variables $X_k$’s are independent of the residuals ($U_{0j}$ and $R_{ij}$).
- The residuals are independent of each other.
Level 1, Case 1: Estimation

To get an estimate of the expected value of $Y_{ij}$, fit the model without any predictors; that is,

$$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij},$$

which is our null/empty model and obtain our estimates $\hat{\gamma}_{00}$, $\hat{\tau}_{0}^{*2}$ and $\hat{\sigma}^{*2}$.

The estimated mean squared error of prediction equals

$$\text{var}(Y_{ij}) = \hat{\tau}_{0}^{*2} + \hat{\sigma}^{*2}.$$
Level 1, Case 2

When the predictors are known, the best guess for $Y_{ij}$ using $X_{k,ij} = x_{k,ij}$.

$$E(Y_{ij}|X_{k,ij} = x_{k,ij}) = E\left[\sum_{k=0}^{p} \gamma_{k}x_{k,ij} + U_{0j} + R_{ij}\right]$$

$$= \sum_{k=0}^{p} \gamma_{k}x_{k,ij} + E[U_{0j}] + E[R_{ij}]$$

$$= \sum_{k=0}^{p} \gamma_{k}0x_{k,ij}$$

...and get $\hat{\tau}^2_0$ and $\hat{\sigma}^2$. 
Level 1, Case 2: Estimation

The estimated mean squared error of prediction,

\[
\frac{1}{n^+} \sum_{j=1}^{N} \sum_{i=1}^{n_j} (Y_{ij} - \hat{Y}_{ij})^2 = \text{var}(Y_{ij} - \hat{Y}_{ij})
\]

\[
= \text{var}(Y_{ij} - \sum_{k=0}^{p} \hat{\gamma}_k x_{k,ij})
\]

\[
= \hat{\tau}_0^2 + \hat{\sigma}^2
\]
The Level 1 Measure

\[ R_1^2 = \frac{\text{var}(Y_{ij}) - \text{var}(Y_{ij} - \sum_{k=0}^{p} \gamma_k x_{k,ij})}{\text{var}(Y_{ij})} \]

\[ = 1 - \frac{\text{var}(Y_{ij} - \sum_{k=0}^{p} \gamma_k x_{k,ij})}{\text{var}(Y_{ij})} \]

\[ = 1 - \frac{\tau_0^2 + \sigma^2}{\tau_0^*2 + \sigma^*2} \]

where \( \tau_0^2 + \sigma^2 \) is from the one with predictor variables and \( \tau_0^*2 + \sigma^*2 \) is from the null model (without) predictor variables.
### HSB Example

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\tau}_0^2$</th>
<th>$\hat{\sigma}^2$</th>
<th>$\hat{\tau}_0^2 + \hat{\sigma}^2$</th>
<th>$R_1^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) null</td>
<td>8.55</td>
<td>39.15</td>
<td>47.70</td>
<td></td>
</tr>
<tr>
<td>(b) cSES</td>
<td>8.61</td>
<td>37.01</td>
<td>45.61</td>
<td>.044</td>
</tr>
<tr>
<td>(c) (b) + minority</td>
<td>6.64</td>
<td>36.12</td>
<td>42.77</td>
<td>.103</td>
</tr>
<tr>
<td>(d) (c) + female</td>
<td>6.26</td>
<td>35.88</td>
<td>42.14</td>
<td>.117</td>
</tr>
<tr>
<td>(e) (d) + sector, size, &amp; meanSES</td>
<td>1.61</td>
<td>35.89</td>
<td>37.50</td>
<td>.214</td>
</tr>
<tr>
<td>(h) (e) + cSES(^\ast)sector female(^\ast)size, &amp; minority(^\ast)sector</td>
<td>1.67</td>
<td>35.59</td>
<td>37.26</td>
<td>.219</td>
</tr>
<tr>
<td>(i) (h) + 5 more cross-level</td>
<td>1.63</td>
<td>35.59</td>
<td>37.22</td>
<td>.220</td>
</tr>
</tbody>
</table>
Level 2: $R^2$ Type Measure

Now we consider predictions of the group means of $Y_{ij}$; that is, $\bar{Y}_{+j}$.

The development is similar to that for Level 1, except now the variance of $\bar{Y}_{+j}$ also depends on the group sample sizes.

The Level 2 measure is

$$R^2 = 1 - \frac{\sigma^2 / n + \tau^2_0}{\sigma^*2 / n + \tau^*_02} \quad \text{with predictors null}$$
$R^2_2$ Type Measure

$$R^2_2 = \frac{\text{var}(\bar{Y}_+j) - \text{var}(\bar{Y}_+j - \sum_{k=1}^{P} \gamma_k 0 \bar{x}_k, +j)}{\text{var}(\bar{Y}_+j)}$$

$$= 1 - \frac{\text{var}(\bar{Y}_+j - \sum_{k=1}^{P} \gamma_k 0 \bar{x}_k, +j)}{\text{var}(\bar{Y}_+j)}$$

$$= 1 - \frac{\sigma^2/n + \tau_0^2}{\sigma^*2/n + \tau^*2}$$

where

- $\sigma^2$ and $\tau_0^2$ are from the model with predictor variables.
- $\sigma^*2$ and $\tau^*2$ are from the null model.
- $n$ = representative value for group sample size.
Representative Sample Size

If the group sample sizes $n_j$ are different, then use either

- A typical value (e.g., if groups are classes and most classes have 25 students).
- The harmonic mean of the sample sizes:

$$\bar{n}_+ = \frac{N}{\sum_{j=1}^{N} (1/n_j)}$$

where

- $N$ is the number of macro units.
- $n_j$ is the number of cases within macro unit $j$. 
HSB example

The harmonic mean equals 41.0587.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\tau}_0^2$</th>
<th>$\hat{\sigma}^2$</th>
<th>$R_1^2$</th>
<th>$R_2^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) null</td>
<td>8.5490</td>
<td>39.1488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) cSES</td>
<td>8.6071</td>
<td>37.0056</td>
<td>.044</td>
<td>.002</td>
</tr>
<tr>
<td>(c) (b) + minority</td>
<td>6.6432</td>
<td>36.1218</td>
<td>.103</td>
<td>.208</td>
</tr>
<tr>
<td>(d) (c) + female</td>
<td>6.2605</td>
<td>35.8773</td>
<td>.117</td>
<td>.249</td>
</tr>
<tr>
<td>(e) (d) + sector, size, meanSES</td>
<td>1.6100</td>
<td>35.8886</td>
<td>.214</td>
<td>.739</td>
</tr>
<tr>
<td>(h) (e) + cSES<em>sector, female</em>size minority*sector</td>
<td>1.6672</td>
<td>35.5940</td>
<td>.219</td>
<td>.733</td>
</tr>
<tr>
<td>(i) (h) + five more cross-level</td>
<td>1.6299</td>
<td>35.5894</td>
<td>.220</td>
<td>.737</td>
</tr>
</tbody>
</table>
$R^2$s for Random Intercept and Slope Models

- The concept is the same; however, with the random effects (slopes), the variances are not constant. Estimation of $R_1^2$ and $R_2^2$ is a bit harder.

- Thanks to a former student, we’ll use the SAS Macro “HLMRSQ.sas” to compute $R_1^2$ (and $R_2^2$) for random slope models.


- If you use this MACRO, use the reference above and the date when the macro was downloaded.
HSB example

Harmonic mean: \( n = 41.0587 \)

<table>
<thead>
<tr>
<th>Model</th>
<th>( \hat{\tau}^2 )</th>
<th>( \hat{\sigma}^2 )</th>
<th>( \hat{\tau}^2 + \hat{\sigma}^2 )</th>
<th>( R_1^2 )</th>
<th>( R_2^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random Intercept Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>null</td>
<td>8.55</td>
<td>39.15</td>
<td>47.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\text{SES}_{ij} - \bar{\text{SES}}_j))</td>
<td>8.61</td>
<td>37.01</td>
<td>45.61</td>
<td>.04</td>
<td>-.00</td>
</tr>
<tr>
<td>+(\bar{\text{SES}}_j)</td>
<td>2.64</td>
<td>37.02</td>
<td>45.61</td>
<td>.17</td>
<td>.63</td>
</tr>
<tr>
<td>+(\text{Female}<em>{ij})+(\text{Minority}</em>{ij})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+(\text{pAcademic}_j+\text{Sector}_j)</td>
<td>1.34</td>
<td>35.88</td>
<td>37.24</td>
<td>.22</td>
<td>.77</td>
</tr>
<tr>
<td><strong>Random Slope/Effects Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((\text{SES}_{ij} - \bar{\text{SES}}_j) + \bar{\text{SES}}_j)</td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
<td>.63</td>
</tr>
<tr>
<td>+(\text{Female}<em>{ij})+(\text{Minority}</em>{ij})</td>
<td></td>
<td></td>
<td></td>
<td>.22</td>
<td>.74</td>
</tr>
<tr>
<td>+(\text{pAcademic}_j+\text{Sector}_j)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+(\text{Minority}_{ij}) (random intercept dropped)</td>
<td></td>
<td></td>
<td></td>
<td>.22</td>
<td>.74</td>
</tr>
</tbody>
</table>

Note: effects in purple are random.
I Using HLMRSQ.sas

1. Download the sas macro: hlmrsq.sas

2. Before using, either
   - Put the marco in a SAS program editor window and “run”, or
   - Add the following to your SAS program
     % include 'C:\... path to...\ hlmrsq.sas';

3. Add the statements (in red) to PROC MIXED code:
   
   ```sas
   proc mixed data=hsbcent noclprint covtest
      method=ML namelen=200;
   class id;
   model mathach = cSES female minority meanSES size sector / solution ;
   random intercept female minority cSES / subject=id
      type=un g ;
   ods output CovParms=cov G=gmat ModelInfo=mod
   SolutionF=solf;
   ```
Using HLMRSQ.sas (continued)

The last statement in SAS program window should be typed exactly as:

```sas
ods output CovParms=cov G=gmat ModelInfo=mod SolutionF=solf;
```

- `ods`, `output`, `CovParms`, `G`, `ModelInfo`, and `SolutionF` are SAS names.
- `cov`, `gmat`, `mod`, and `solf` are names given to these things and are the names using in the SAS marco.

4 The following command will execute the macro:

```sas
%hlmrsq(CovParms=cov,GMatrix=gmat,ModelInfo=mod,SolutionF=solf);
```

5 Output:

<table>
<thead>
<tr>
<th>Explained Proportion of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep Size</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>41.06</td>
</tr>
</tbody>
</table>
R and $R^2$s

- I could not find a function or package that does this.
- I wrote a function called hlmRsq.
- To use it,
  - Run the function.
  - Fit model: `model1 <- lmer(mathach ~ 1 + cses + meanSES + (1 + cses | id ), data=hsb, REML=FALSE)`
    - **If you have random slope, put them in model first.**
  - `hlmRsq(hsb,model1,mathach,id)`

**Results:**

```
hlmRsq(hsb,model1)
```

<table>
<thead>
<tr>
<th>harmonic.mean</th>
<th>R1sq</th>
<th>R2sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>41.05874</td>
<td>0.1697781</td>
<td>0.6290596</td>
</tr>
</tbody>
</table>
Notes: $R^2$ Measures for HLMs

- In the **population**, $R^1_1$ and $R^2_2$ never decrease when you add explanatory variables, but they can in the sample.

- A large decrease or negative values of $R^1_1$ and/or $R^2_2$ may indicate a misspecified model for the fixed effects. In particular, the problem may be that you’ve made an (implicit) restriction such that a variable’s within-group and between group-coefficients are the same, but in the population they differ.

In our example, there is a negative value for $R^2_2 (=-.0006)$. This could result from the model being too simple.

In the HSB example, the Level 1 and Level 2 effects of student SES are different; however, in the model that included $SES_{ij}$ as the only predictor implicitly restricts these effects to be equal. In this case

$$R^2_2 = 1 - \frac{8.61 + 37.01/41.0587}{8.55 + 39.15/41.06} = -.0008$$
Why $R^2$'s are Similar?

The $R^2$'s for random slope models should be (and generally are) very similar in value to those from the random intercept models.

In a random slope model,

$$Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + U_{0j} + U_{1j}x_{ij} + R_{ij}$$

On average, $E(U_j) = 0$, and $X_{ij}$, $U_j$ and $R_{ij}$ are independent of each other.

One Final Comment: Like the information criteria, $R^2_1$ and $R^2_2$ are indices of fit and are not statistical significance tests.
Summary

Tests for fixed effects:
- Wald, $t$ and $F$ tests $\rightarrow$ OK under MLE and REML.
- Likelihood ratio only valid under MLE.

Test for random effects:
- Testing $H_0: \tau^2 = 0$ is a non-standard test.
- Normality assumption required for $z$ (Wald) test completely fails.
- A Regularity condition for valid likelihood ratio test is not met.
- Can compute likelihood ratio test statistic for $q$ versus $q + 1$ random effects where the sampling distribution of the test statistics follows a mixture of $\chi^2_{q+1}$ and $\chi^2_q$.

Global measures:
- Information criteria: useful for model comparison.
- $R^2_1$ and $R^2_2$: can detect model miss-specification.