Statistical Inference: The Marginal Model

Edps/Psych/Soc 587

Carolyn J. Anderson

Department of Educational Psychology

ILLINOIS

Outline

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

Reading: Snijders & Bosker, Chapter 6

Additional References

- Verbeke,G, & Molenbergs, G (2000). *Linear Mixed Models for Longitudinal Data*. NY: Springer.
- Snijders,T.A.B., & Bosker, R.J. (1994). Modelled variance in two-level models. *Sociological Methods & Research, 22,* 342–363.
- Self, S.G., & Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood tests under nonstandard conditions. *Journal of the American Statistical Association, 82*, 605–610.
- Goldstein, H. (1999). *Multilevel statistical models*.
- Stram, D.0., & Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50*, 1171–1177.
- Stram, D.O., & Lee, J.W. (1995). Correction to: Variance components testing in the longitudinal mixed effects model. *Biometrics, 51*, 1196.

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, $\hat{\mathbf{\Gamma}} = (\hat{\gamma}_{00}, \hat{\gamma}_{01}, \ldots)'$.

The asymptotic sampling distribution of $\hat{\mathbf{\Gamma}}$ is

$$\hat{\mathbf{\Gamma}} \sim \mathcal{N}\left(\mathbf{\Gamma}, \text{cov}(\hat{\mathbf{\Gamma}})\right) \quad \text{where} \quad \hat{\mathbf{\Gamma}} = \left(\sum_{j=1}^{N} \boldsymbol{X}_j' \hat{\boldsymbol{V}}_j^{-1} \boldsymbol{X}_j\right)^{-1} \sum_{j=1}^{N} \boldsymbol{X}_j' \hat{\boldsymbol{V}}_j^{-1} \boldsymbol{y}_j$$

where $\hat{\boldsymbol{V}}_j$ is the estimated covariance matrix of $\boldsymbol{Y}_j$, which equals

$$\hat{\boldsymbol{V}}_j = \boldsymbol{Z}_j \hat{\boldsymbol{T}} \boldsymbol{Z}_j' + \hat{\sigma}^2 \boldsymbol{I}$$

Our estimate of $\mathbf{\Gamma}$ depends on $\hat{\boldsymbol{T}}$ and $\hat{\sigma}^2$.

Covariance Matrix of $\hat{\Gamma}$

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

IF

- The model for the mean of Y_j is correctly specified, (i.e., $\boldsymbol{X}_j\boldsymbol{\Gamma}$) so $\text{E}(\hat{\boldsymbol{\Gamma}}) = \boldsymbol{\Gamma}$ (i.e, unbiased).
- The marginal covariance matrix is correctly specified, (i.e., $\boldsymbol{V}_j = \boldsymbol{Z}_j\boldsymbol{T}\boldsymbol{Z}_j' + \sigma^2\boldsymbol{I}$) so the covariance matrix of data equals the predicted covariance matrix.

Covariance Matrix of $\hat{\Gamma}$

THEN

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

where

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} \quad \text{and} \quad \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{V}_N \end{pmatrix}$$

We can now use the fact $\hat{\boldsymbol{\Gamma}} \sim \mathcal{N}\left(\boldsymbol{\Gamma}, \text{cov}(\hat{\boldsymbol{\Gamma}})\right)$

Digression: Distribution of $\hat{\Gamma}$

- Since $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{X\Gamma}, \boldsymbol{\Sigma_Y})$ and

$$\begin{aligned}\hat{\boldsymbol{\Gamma}} &= \left(\sum_{j=1}^{N} \boldsymbol{X}_j' \hat{\boldsymbol{V}}_j^{-1} \boldsymbol{X}_j\right)^{-1} \sum_{j=1}^{N} \boldsymbol{X}_j' \hat{\boldsymbol{V}}_j^{-1} \boldsymbol{y}_j \\ &= \underbrace{(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}}\boldsymbol{y} \\ &= \qquad \boldsymbol{A} \qquad\qquad \boldsymbol{y}\end{aligned}$$

- The Expected value,

$$\begin{aligned}\mathsf{E}(\hat{\boldsymbol{\Gamma}}) &= \mathsf{E}[(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1})\boldsymbol{Y}] \\ &= (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\mathsf{E}[(\boldsymbol{X\Gamma} + \boldsymbol{\epsilon})] \\ &= (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})\mathsf{E}[(\boldsymbol{\Gamma} + \boldsymbol{\epsilon})] \\ &= \boldsymbol{\Gamma}\end{aligned}$$

Distribution of $\hat{\Gamma}$ (continued)

- Covariance matrix,

$$
\begin{aligned}
\text{cov}(\hat{\boldsymbol{\Gamma}}) &= \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}' \\
&= [(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}]\underbrace{\overbrace{\boldsymbol{\Sigma}_{\boldsymbol{Y}}}^{\boldsymbol{V}}[\boldsymbol{V}^{-1}}_{\boldsymbol{I}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}] \\
&= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\underbrace{\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}}_{\boldsymbol{I}} \\
&= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}
\end{aligned}
$$

- Since $\hat{\boldsymbol{\Gamma}}$ is a linear function of a vector of normal random variables (i.e., $\boldsymbol{Y}$), $\hat{\boldsymbol{\Gamma}}$ is normal.
- So $\hat{\boldsymbol{\Gamma}} \sim \mathcal{N}(\boldsymbol{\Gamma}, (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1})$

Approximate Wald Tests

- Perform statistical hypothesis tests on
  - One γ, e.g.,
  $$H_o : \gamma_{01} = 0 \text{ versus } H_o : \gamma_{01} \neq 0$$
  - Multiple γ's, including contrasts, e.g.,
  $$H_o : \boldsymbol{L\Gamma} = \mathbf{0} \text{ versus } H_a : \boldsymbol{L\Gamma} \neq \mathbf{0}$$
- Form confidence intervals for parameters.

One Fixed Effect

Sampling distribution for one fixed effect,

$$\hat{\gamma}_{kl} \sim \mathcal{N}(\gamma_{kl}, \text{var}(\hat{\gamma}_{kl}))$$

Statistical Hypothesis:

$$H_o : \gamma_{kl} = \gamma^*_{kl} \qquad \text{versus} \qquad H_a : \gamma_{kl} \neq \gamma^*_{kl}.$$

Note:

- Usually, $\gamma^*_{kl} = 0$
- Can do directional tests, i.e.,

$$H_a : \gamma_{kl} > \gamma^*_{kl} \qquad \text{or} \qquad H_a : \gamma_{kl} < \gamma^*_{kl}$$

Test statistic and approximate sampling distribution:

$$z = \frac{\hat{\gamma}_{kl} - \gamma^*_{kl}}{\widehat{SE}} \sim \mathcal{N}(0,1) \qquad \text{or} \qquad z^2 \sim \chi^2_1$$

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:

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\overline{\text{SES}})_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{sector})_j + \gamma_{12}(\text{size})_j + U_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}(\text{sector})_j + \gamma_{22}(\text{size})_j + \gamma_{23}(\overline{\text{SES}})_j + U_{2j} \\
\beta_{3j} &= \gamma_{30} + \gamma_{31}(\text{sector})_j + \gamma_{32}(\text{size})_j + \gamma_{33}(\overline{\text{SES}})_j + U_{3j}
\end{aligned}$$

HSB: Linear Mixed Model

$$
\begin{aligned}
(\text{math})_{ij} &= [\gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\overline{\text{SES}})_j] \\
&\quad +[\gamma_{10} + \gamma_{11}(\text{sector})_j + \gamma_{12}(\text{size})_j](\text{cSES})_{ij} \\
&\quad +[\gamma_{20} + \gamma_{21}(\text{sector})_j + \gamma_{22}(\text{size})_j + \gamma_{23}(\overline{\text{SES}})_j](\text{female})_{ij} \\
&\quad +[\gamma_{30} + \gamma_{31}(\text{sector})_j + \gamma_{32}(\text{size})_j + \gamma_{33}(\overline{\text{SES}})_j](\text{minority})_{ij} \\
&\quad +U_{0j} + U_{ij}(\text{cSES})_{ij} + U_{2j}(\text{female})_{ij} + U_{3j}(\text{minority})_{ij} + R_{ij} \\
&= \gamma_{00} + \gamma_{10}(\text{cSES})_{ij} + \gamma_{20}(\text{female})_{ij} + \gamma_{30}(\text{minority})_{ij} \\
&\quad +\gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\overline{\text{SES}})_j \\
&\quad +\gamma_{11}(\text{sector})_j(\text{cSES})_{ij} + \gamma_{12}(\text{size})_j(\text{cSES})_{ij} \\
&\quad +\gamma_{21}(\text{sector})_j(\text{female})_{ij} + \gamma_{22}(\text{size})_j(\text{female})_{ij} \\
&\quad +\gamma_{23}(\overline{\text{SES}})_j(\text{female})_{ij} + \gamma_{31}(\text{sector})_j(\text{minority})_{ij} \\
&\quad +\gamma_{32}(\text{size})_j(\text{minority})_{ij} + \gamma_{33}(\overline{\text{SES}})_j(\text{minority})_{ij} \\
&\quad +U_{0j} + U_{1j}(\text{cSES})_{ij} + U_{2j}(\text{female})_{ij} + U_{3j}(\text{minority})_{ij} + R_{ij}
\end{aligned}
$$

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

$$H_o : \gamma_{kl} = 0 \qquad \text{versus} \qquad H_a : \gamma_{kl} \neq 0.$$

- 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 from SAS

Effect	Estimate	se	DF	Wald	p	
Intercept	12.0260	0.4691				
cses	2.2823	0.3111	1	53.81	$< .0001$	*
female	−0.3402	0.4637	1	0.54	0.4632	
minority	−4.2068	0.6714	1	39.26	$< .0001$	*
meanses	4.2207	0.5003	1	71.17	$< .0001$	*
size	0.001125	.000314	1	12.86	0.0003	*
sector	1.7360	0.4173	1	17.31	$< .0001$	*
cses*size	0.000032	.000203	1	0.03	0.8728	
cses*sector	−1.0033	0.2528	1	15.74	$< .0001$	*
female*meanses	−0.03207	0.4838	1	0.00	0.9471	
female*size	−0.00070	.000304	1	5.36	0.0206	*
female*sector	−0.3006	0.4284	1	0.49	0.4829	
minority*meanses	−0.7793	0.5391	1	2.09	0.1483	
minority*size	0.000183	.000398	1	0.21	0.6446	
minority*sector	2.1189	0.5430	1	15.23	$< .0001$	*

R & Wald Tests

We just use the fact that

$$\text{Wald} = t_1^2 = \frac{\hat{\gamma_{jk}}^2}{\text{var}(\hat{\gamma}_{jk})} \sim \chi_1^2$$

In the output (if you're using lmerTest), you will get t, so just square this.

```
s4 ← summary(model4)
s4 ← as.data.frame(s4[10])
names(s4) ← c("Estimate","StdError","df t","t","Pr(> |t|)")
s4$df.Wald ← rep(1,nrow(s4))
s4$Wald4 ← s4$t**2
```

R & Wald Tests

For output that looks reasonable (for the most part), use the `xtable` package and

```
options(scipen = 999)    # Turns off scientific notationn
print(s4, type = ‘‘html’’,digits=2)
options(scipen = 0)    # Turns scientific notation back on
```

Note `print` is quirky. digits=2 actually gave me 3 digits.

R & Wald Tests

	Estimate	StdError	df	Wald	p
(Intercept)	12.026	0.469			
female	-0.340	0.464	1.000	0.538	.465
minority	-4.206	0.671	1.000	39.252	0.000
cses	2.282	0.311	1.000	53.793	0.000
meanses	4.220	0.500	1.000	71.139	0.000
zsize	0.680	0.190	1.000	12.859	0.000
sector	1.736	0.417	1.000	17.304	0.000
cses:zsize	0.020	0.123	1.000	0.026	0.873
cses:sector	-1.003	0.253	1.000	15.739	0.000
female:meanses	-0.031	0.484	1.000	0.004	0.948
female:zsize	-0.425	0.184	1.000	5.357	0.022
female:sector	-0.301	0.428	1.000	0.492	0.484
minority:meanses	-0.779	0.539	1.000	2.091	0.151
minority:zsize	0.111	0.240	1.000	0.212	0.646
minority:sector	2.119	0.543	1.000	15.231	0.000

Confidence Intervals for γ_{kl}'s

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

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

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

$$2.2823 \pm 1.96(0.3111) \longrightarrow (1.67, 2.89)$$

R & Wald Confidence Intervals for γ_{kl}'s

Since we have s4, the summary of model 4 as an object, we can use information to compute confidence intervals. Below is code for 95% intervals

```
names(s4)                    # check names of things
s4$upper ← s4$Estimate - qnorm(.025)*s4$StdError
s4$lower ← s4$Estimate - qnorm(.975)*s4$StdError
s4
round(s4[,8:9],digits=2)
```

Note: Later we'll look at methods that use alternative methods to estimate confidence intervals

R & Wald confidence Intervals for γ_{kl}'s from R

	lower	upper
(Intercept)	11.11	12.95
female	-1.25	0.57
minority	-5.52	-2.89
cses	1.67	2.89
meanses	3.24	5.20
zsize	0.31	1.05
sector	0.92	2.55
cses:zsize	-0.22	0.26
cses:sector	-1.50	-0.51
female:meanses	-0.98	0.92
female:zsize	-0.79	-0.07
female:sector	-1.14	0.54
minority:meanses	-1.84	0.28
minority:zsize	-0.36	0.58
minority:sector	1.05	3.18

General Tests on Fixed Effects

We may want to

- Simultaneously test a set of γ'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 in SAS or as.factor( ) in R which create dummy codes).
- One or more contrasts of γ's (e.g., to test whether some γ's are equal).

General Tests on Fixed Effects

For the general case, tests are based on the fact

$$\hat{\boldsymbol{\Gamma}} \sim \mathcal{N}\left(\boldsymbol{\Gamma}, \text{cov}(\hat{\boldsymbol{\Gamma}})\right)$$

Hypotheses are in the form of

$$H_o : \boldsymbol{L\Gamma} = \mathbf{0} \qquad \text{versus} \qquad H_a : \boldsymbol{L\Gamma} \neq \mathbf{0}$$

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

In Scaler From:

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

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

General Test Statistic

$$\hat{\boldsymbol{\Gamma}}'\boldsymbol{L}'\underbrace{\left[\boldsymbol{L}\left(\sum_{j=1}^{N}\boldsymbol{X}_j'\hat{\boldsymbol{V}}_j^{-1}\boldsymbol{X}_j\right)^{-1}\boldsymbol{L}'\right]^{-1}}_{\text{covariance matrix of } \boldsymbol{L}\hat{\boldsymbol{\Gamma}}}\boldsymbol{L}\hat{\boldsymbol{\Gamma}}$$

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

I won't make you compute this by hand. . . Let SAS or R do the busy-work. In R, use the function `contrast` that I wrote.

HSB: General Test Statistic

In our example $\mathbf{\Gamma}$ is a (15×1) vector:

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

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

Interaction	Parameter	Interaction	Parameter
$(\text{size})_j(\text{cSES})_{ij}$	γ_{12}	$(\text{size})_j(\text{minority})_{ij}$	γ_{32}
$(\text{sector})_j(\text{female})_{ij}$	γ_{21}	$(\overline{\text{SES}})_j(\text{minority})_{ij}$	γ_{33}
$(\overline{\text{SES}})_j(\text{female})_{ij}$	γ_{23}		

Simultaneously Testing γ_{rk}'s

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

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

$$\boldsymbol{L} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Simultaneously Testing Cross-Level

Statistical hypotheses are

$$H_o : \boldsymbol{L\Gamma} = \begin{pmatrix} \gamma_{12} \\ \gamma_{21} \\ \gamma_{23} \\ \gamma_{32} \\ \gamma_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{vs} \quad H_a : \boldsymbol{L\Gamma} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

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:

Contrasts

Label	DF	Chi-Square	Pr>ChiSq
Cross-level interactions	5	2.97	0.7048

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 $\boldsymbol{L}$.
  - Output includes $\boldsymbol{L}\hat{\boldsymbol{\Gamma}}$ and it's the S.E. of $\boldsymbol{L}\hat{\boldsymbol{\Gamma}}$, 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, "contrast")
- Include a `source` command, e.g., `source("All_.txt")`
- Create $\boldsymbol{L}$ that has rows as tests/constrasts and columns correspond to fixed effects. Talk about requirements for $\boldsymbol{L}$ in class.
- `contrast(model, L)`
- Returns table with F, numerator df, a guess at denomonator 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

```
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
round(contrast(cmodel, L), digits=2)
    F   num df   den df   p-value     X2     df   p-chisquare
 1.73     5.00   156.14      0.13   8.66   5.00          0.12
```

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 the estimated standard errors are too small $\longrightarrow$ 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 γ, use Students t–distribution instead of the standard normal.

The test statistic is still

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

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

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

which is compared to an $\mathcal{F}$ distribution where the numerator degrees of freedom equals c (i.e., rank of $\boldsymbol{L}$, number of tests/contrasts performed; that is, the number of rows in $\boldsymbol{L}$). 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 $n_+ -$ (number of parameters estimated).
- ddfm=betwithin. This is the default when you have a REPEATED statement and recommended instead of contain when the $\boldsymbol{Z}_j$ 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	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

Effect	Num DF	Den DF	Chi-Square	F Value	Pr > $ChiSq$	Pr> F
cses	1	148	53.81	53.81	$< .0001$	$< .0001$
female	1	123	0.54	0.54	0.4632	0.4646
minority	1	157	39.26	39.26	$< .0001$	$< .0001$
meanses	1	182	71.17	71.17	$< .0001$	$< .0001$
size	1	156	12.86	12.86	0.0003	0.0004
sector	1	134	17.31	17.31	$< .0001$	$< .0001$
cses*size	1	155	0.03	0.03	0.8728	0.8731
cses*sector	1	148	15.74	15.74	$< .0001$	0.0001
female*meanses	1	166	0.00	0.00	0.9471	0.9472
female*size	1	133	5.36	5.36	0.0206	0.0222
female*sector	1	143	0.49	0.49	0.4829	0.4840
minority*meanses	1	120	2.09	2.09	0.1483	0.1509
minority*size	1	142	0.21	0.21	0.6446	0.6453
minority*sector	1	133	15.23	15.23	$< .0001$	0.0002

Output: 99% Confidence Limits

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

Effect	Lower	Upper
cses	1.4710	3.0936
female	−1.5537	0.8734
minority	−5.9614	−2.4521
meanses	2.9316	5.5098
size	0.000317	0.001933
sector	0.6608	2.8112
cses*size	−0.00049	0.000555
cses*sector	−1.6548	−0.3518
female*meanses	−1.2785	1.2144
female*size	−0.00149	0.000080
female*sector	−1.4044	0.8032
minority*meanses	−2.1683	0.6098
minority*size	−0.00084	0.001208
minority*sector	0.7199	3.5179

For R Users

- Use the `lmerTest` package. The lmerTest pckage gives Sattherwaite 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 can take a very long time for complex models. I illustrate it using a simpler one

```
model1 <- lmer(matach ~ 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

		2.5 %	97.5 %
$\sqrt{\tau_0^2}$	.sig01	1.0155	1.4559
$\sqrt{\sigma^2}$	.sigma	5.8930	6.0834
	(Intercept)	11.7390	13.1435
	cses	1.7063	2.1080
	female	-1.5455	-0.9280
	minority	-3.2955	-2.4970
	meanses	3.2806	4.6013
	zsize	0.1380	0.6793
	sector	1.5392	2.7345

Alternatively, you can use profile likelihood to get confidence intervals, which doesn't take as long:

```
profile.ci ← confint(model1, nsim=1000, level=0.99)
round(profile.ci, digits=4)
```

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., $\boldsymbol{X}_j\boldsymbol{\Gamma}$) so that $\text{E}(\hat{\boldsymbol{\Gamma}}) = \boldsymbol{\Gamma}$ (i.e, unbiased).
  - The marginal covariance matrix is correctly specified, (i.e., $\boldsymbol{V}_j = \boldsymbol{Z}_j\boldsymbol{T}\boldsymbol{Z}_j' + \sigma^2\boldsymbol{I}$) 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 $= \boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\Gamma}}$) rather than the covariance matrices of the data (i.e., the $\boldsymbol{Y}_j$'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 Specially What It Is

Recall (page 9),

$$\begin{aligned}\text{cov}(\hat{\boldsymbol{\Gamma}}) &= \left[(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\right]\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{\Sigma_Y}\boldsymbol{V}^{-1}\boldsymbol{X}\left[(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\right] \\ &= M'\boldsymbol{\Sigma_Y}M\end{aligned}$$

- Replace $\boldsymbol{\Sigma_Y}$ with
$$(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\Gamma}})(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\Gamma}})',$$
which is a block diagonal matrix with $(\boldsymbol{y}_j-\boldsymbol{X}_j\hat{\boldsymbol{\Gamma}})(\boldsymbol{y}_j-\boldsymbol{X}_j\hat{\boldsymbol{\Gamma}})'$ 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

	Correct Model			
		model based		sandwich
Parm	est	se	est	se
Random effects				
τ_{00}	0.8201		0.8201	
τ_{10}	-0.1018		-0.1018	
τ_{11}	0.8581		0.8581	
τ_{20}	-0.1563		-0.1563	
τ_{21}	0.1157		0.1157	
τ_{22}	1.1409		1.1409	
σ^2	3.9115		3.9115	
Fixed effects				
γ_{00}	5.1656	0.1141	5.1656	0.1141
γ_{10}	2.0524	0.0960	2.0524	0.0960
γ_{20}	3.0058	0.1086	3.0058	0.1086

Simulation

	Correct Model		Wrong Model			
		model based		model based		sandwich
Effect	est	se	est	se	est	se
Random effects						
τ_{00}	0.8201		0.5352		0.5352	
τ_{10}	-0.1018		-0.0477		-0.0477	
τ_{11}	0.8581		0.8055		0.8055	
τ_{20}	-0.1563					
τ_{21}	0.1157					
τ_{22}	1.1409					
σ^2	3.9115		21.0821		21.0821	
Fixed effects						
γ_{00}	5.1656	0.1141	5.0903	0.1686	5.0903	0.1668
γ_{10}	2.0524	0.0960	2.0112	0.1035	2.0112	0.1035
γ_{20}	3.0058	0.1086	3.1108	0.0387	3.1108	0.1148

Simulation

	Correct Model		Even Worse Model			
		model based		model based		sandwich
Effect	est	se	est	se	est	se
Random effects						
τ_{00}	0.8201		1.3180		1.3180	
τ_{10}	-0.1018					
τ_{11}	0.8581					
τ_{20}	-0.1563					
τ_{21}	0.1157					
τ_{22}	1.1409					
σ^2	3.9115		28.3097		28.3097	
Fixed effects						
γ_{00}	5.1656	0.1141	5.1555	0.2037	5.1555	0.2034
γ_{10}	2.0524	0.0960	2.0177	0.0559	2.0177	0.1150
γ_{20}	3.0058	0.1086	3.1070	0.0432	3.1070	0.1134

Eg. of Robust/Empirical Estimation

Specify the "empirical" option in the PROC MIXED statement.

```
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	$\longleftarrow$ changed
Degrees of Freedom Method	Containment	

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t 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

Label	Num DF	Den DF	Chi-Square	F Value	Pr>ChiSq	Pr> F
Cross-level interactions	5	6604	3.32	0.66	.6503	.6503

Comparison: Model versus Robust

	Robust Estimation		Model-Based	
Effect	Estimate	Std. Error	Estimate	Std. Error
Intercept	12.0260	0.4269	12.0260	0.4691
cses	2.2823	0.3176	2.2823	0.3111
female	−0.3402	0.4137	−0.3402	0.4637
minority	−4.2068	0.6439	−4.2068	0.6714
meanses	4.2207	0.4961	4.2207	0.5003
size	0.001125	0.000296	0.001125	0.000314
sector	1.7360	0.3978	1.7360	0.4173
cses*size	0.000032	0.000222	0.000032	0.000203
cses*sector	−1.0033	0.2503	−1.0033	0.2528
female*meanses	−0.03207	0.4235	−0.03207	0.4838
female*size	−0.00070	0.000256	−0.00070	0.000304
female*sector	−0.3006	0.4150	−0.3006	0.4284
minority*meanses	−0.7793	0.4933	−0.7793	0.5391
minority*size	0.000183	0.000386	0.000183	0.000398
minority*sector	2.1189	0.5398	2.1189	0.5430

For R Users: Empirical SEs

- `lmer` does not compute these, so I wrote a function to compute these: "robust.txt". Later I found some code online, but that I found online but it is not working.
- Use the `robust` to will compute them:
  - `source(''All_functions.txt'')`
  - Fit a model, say model 3

    ```
    summary(model3 <- lmer(mathach ~ 1 + cses + female +
    minority + meanses + zsize + sector + cses*zsize +
    cses*sector + female*meanses + female*zsize +
    female*sector + minority*meanses + minority*zsize +
    minority*sector + (1 + cses | id), data=hsb, REML=FALSE))
    ```
  - To get robust/sandwich standard errors, type

    ```
    r3 <- robust(model3, hsb$mathach, hsb$id,
    "between/within")
    round(r3, digits=4)
    ```

r3

Fixed	between/ /within	Model se.	Model t	p	Robust se	Robust t	p
(Intercept)	155.00	0.43	27.78	0.00	0.42	28.55	0.00
cses	7015.00	0.31	7.25	0.00	0.32	7.07	0.00
female	7015.00	0.42	-0.85	0.40	0.41	-0.88	0.38
minority	7015.00	0.62	-6.61	0.00	0.65	-6.29	0.00
meanses	155.00	0.46	9.15	0.00	0.49	8.58	0.00
zsize	155.00	0.18	3.94	0.00	0.18	3.88	0.00
sector	155.00	0.38	4.57	0.00	0.39	4.45	0.00
cses:zsize	7015.00	0.12	0.26	0.79	0.13	0.24	0.81
cses:sector	155.00	0.25	-3.93	0.00	0.25	-4.00	0.00
female:meanses	7015.00	0.44	-0.12	0.90	0.42	-0.13	0.90
female:zsize	7015.00	0.17	-2.48	0.01	0.15	-2.75	0.01
female:sector	155.00	0.39	-0.60	0.55	0.41	-0.58	0.56
minority:meanses	7015.00	0.49	-1.64	0.10	0.49	-1.64	0.10
minority:zsize	7015.00	0.22	0.31	0.75	0.23	0.30	0.77
minority:sector	155.00	0.49	4.18	0.00	0.54	3.83	0.00

Comparison: Model versus Robust

Notes:

- Estimates of fixed effect are exactly the same (not shown) and will always be 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

$$H_o : \gamma_{kl} = 0 \quad \text{versus} \quad H_a : \gamma_{kl} \neq 0$$

Likelihood Ratio Test Statistic

is defined as

$$\begin{aligned} -2\ln\lambda_N &= -2\ln\left[\frac{L_{ML}(\hat{\mathbf{\Gamma}}_o, \hat{\boldsymbol{T}}, \hat{\sigma}^2)}{L_{ML}(\hat{\mathbf{\Gamma}}, \hat{\boldsymbol{T}}, \hat{\sigma}^2)}\right] \\ &= -2(\ln[L_{ML}(\hat{\mathbf{\Gamma}}_o, \hat{\boldsymbol{T}}, \hat{\sigma}^2)] - \ln[L_{ML}(\hat{\mathbf{\Gamma}}, \hat{\boldsymbol{T}}, \hat{\sigma}^2)]) \end{aligned}$$

where

- $L_{ML}(\hat{\mathbf{\Gamma}}_o, \hat{\boldsymbol{T}}, \hat{\sigma}^2) =$ the value of the likelihood function under the nested model.
- $L_{ML}(\hat{\mathbf{\Gamma}}, \hat{\boldsymbol{T}}, \hat{\sigma}^2) =$ 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 χ^2 random variable with degrees of freedom equal to the difference between the number of γ's in the two models.

The likelihood ratio test for fixed effects is only valid for ML estimation.

L.R. Test & Estimation Method

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 $\hat{\boldsymbol{T}}$ & $\hat{\sigma}^2$, use standard estimation techniques to estimate the mean structure (i.e., the γ'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

$$H_o : \boldsymbol{L\Gamma} = \begin{pmatrix} \gamma_{12} \\ \gamma_{21} \\ \gamma_{23} \\ \gamma_{32} \\ \gamma_{33} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \begin{matrix} (\text{size})_j(\text{cSES})_{ij} \\ (\text{sector})_j(\text{female})_{ij} \\ (\overline{\text{SES}})_j(\text{female})_{ij} \\ (\text{size})_j(\text{minority})_{ij} \\ (\overline{\text{SES}})_j(\text{minority})_{ij} \end{matrix}$$

versus

$$H_a : \boldsymbol{L\Gamma} \neq \mathbf{0}$$

Example of Likelihood Ratio Test (continued)

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

	Estimation method	
Model	ML	REML
Reduced or null	46,223.5645	46,263.2394
Full	46,220.8436	46,288.5541
$-2\ln\lambda_N =$	2.7209	-25.3147
$df =$	5	
p-value =	.74	

Example of Likelihood Ratio Test (R)

```
> anova(model1,model2)
Data:  hsb
Models:
model1:  mathach ∼ 1 + cses + female + minority + meanses +
sdsize + sector +
model1:  (1 | id)
model2:  mathach ∼ 1 + cses + female + minority + meanses +
sdsize + sector +
model2:  cses * sdsize + (1 | id)
        Df    AIC    BIC  logLik  deviance   Chisq  Chi Df  P
 model1   9  46307  46369  -23145     46289
 model2  10  46302  46370  -23141     46282  7.7638       1  (
 ---
Signif.  codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Summary: Tests for Fixed Effects

Test Statistic	Value	Distribution	p-value
Model Based			
Wald	2.97	χ^2_5	.70
F	.59	$\mathcal{F}_{5,6604}$	.70
$-2\ln\lambda_N$	2.72	χ^2_5	.74
Robust Estimation			
Wald	3.32	χ^2_5	.65
F	.66	$\mathcal{F}_{5,6604}$	.65

Before Tests for Variance Components

Simplify by dropping the 5 cross-level interactions.

Level 1

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

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

Level 2

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{sector})_j + \gamma_{02}(\text{size})_j + \gamma_{03}(\overline{\text{SES}})_j + U_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{sector})_j + U_{1j} \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}(\text{size})_j + U_{2j} \\
\beta_{3j} &= \gamma_{30} + \gamma_{31}(\text{sector})_j + U_{3j}
\end{aligned}$$

Linear Mixed Model

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

Simpler Model: 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

Convergence criteria met.

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t 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...

$$
\begin{aligned}
\widehat{(\text{math})}_{ij} \quad = \quad & [12.084 + 1.618(\text{sector})_j \\
& +.001(\text{size})_j + 3.98(\overline{\text{SES}})_j] \\
& +[2.324 - 1.012(\text{sector})_j](\text{cSES})_{ij} \\
& +[-.540 - .001(\text{size})_j](\text{female})_{ij} \\
& +[-3.793 + 1.765(\text{sector})_j](\text{minority})_{ij}
\end{aligned}
$$

...for now...

Inference for Variance Components

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

- Useful for interpreting random variation in the data.
- Essential for model-based inferences.
  - Over-parameterization of covariance structure $\longrightarrow$ inefficient (and possibly poor) estimated standard errors for the fixed effects.
  - Under-parameterization of covariance structure $\longrightarrow$ 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 $(-\boldsymbol{H})^{-1}$, where $\boldsymbol{H}$ is the Hessian.
- Wald tests (& confidence statements) for:
  1. Variances, i.e.,

$$H_o : \tau_k^2 = 0 \qquad \text{versus} \qquad H_a : \tau_k^2 \neq 0$$

  2. Covariances, e.g.,

$$H_o : \tau_{kl} = 0 \qquad \text{versus} \qquad H_a : \tau_{kl} \neq 0 \qquad \text{for } k \neq l$$

Approximate Wald Tests: Variances

For example: $H_o : \tau_k^2 = 0$ versus $H_a : \tau_k^2 \neq 0$

- The closer τ_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 τ_k^2 cannot be non-negative.

Variance: Wald Test Statistic

$$z = \frac{\hat{\tau}_k^2}{\widehat{S.E.}}$$

Example: HSB data and SAS/MIXED commands:

```
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

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	id	2.2408	0.4991	4.49	$<$ .0001
UN(2,2)	id	0.6791	0.5117	1.33	0.0922
UN(3,3)	id	0.9088	0.6936	1.31	0.0951
UN(4,4)	id	0.1412	0.2118	0.67	0.2525
Residual		35.3169	0.6106	57.84	$<$ .0001

Covariances

- For example,

$$H_o : \tau_{kl} = 0 \quad \text{for} \quad k \neq l \quad \text{versus} \qquad H_a : \tau_{kl} \neq 0$$

- 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 τ_k^2's are greater than 0.
- SAS/MIXED results for covariances (and variances)...

Covariances Parameter Estimates

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

"*" indicates statistics for a variance.

Likelihood Ratio Test for Variances?

The Likelihood ratio test statistic for variance components is

$$\begin{aligned} -2\ln\lambda_N &= -2\ln\left[\frac{L_{ML}(\hat{\mathbf{\Gamma}},\hat{\boldsymbol{T}}_o,\hat{\sigma}^2)}{L_{ML}(\hat{\mathbf{\Gamma}},\hat{\boldsymbol{T}},\hat{\sigma}^2)}\right] \\ &= -2(\ln[L_{ML}(\hat{\mathbf{\Gamma}},\hat{\boldsymbol{T}}_o,\hat{\sigma}^2)] - \ln[L_{ML}(\hat{\mathbf{\Gamma}},\hat{\boldsymbol{T}},\hat{\sigma}^2)], \end{aligned}$$

where

- $L_{ML}(\hat{\mathbf{\Gamma}},\hat{\boldsymbol{T}}_o,\hat{\sigma}^2)$ = the value of the likelihood function under the nest model.
- $L_{ML}(\hat{\mathbf{\Gamma}},\hat{\boldsymbol{T}},\hat{\sigma}^2)$ = 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 χ^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 $H_o : \tau_k^2 = 0$.

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:

$$H_o : \tau_{30} = \tau_{31} = \tau_{32} = \tau_3^2 = 0.$$

- 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

- Self, S.G., & Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood tests under nonstandard conditions. *Journal of the American Statistical Association, 82*, 605–610.
- Stram, D.0., & Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics, 50*, 1171–1177.
- Stram, D.O., & Lee, J.W. (1995). Correction to: Variance components testing in the longitudinal mixed effects model. *Biometrics, 51*, 1196.

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 χ^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:

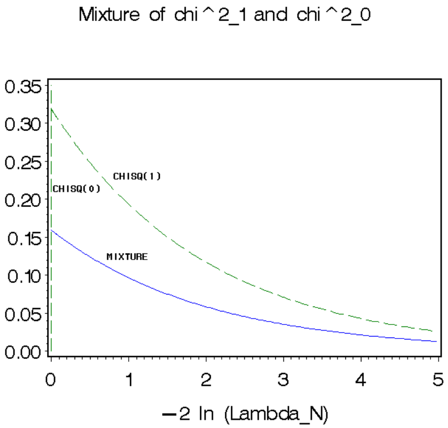
$$H_o : \tau_0^2 = 0 \quad \text{versus} \quad \tau_0^2 > 0$$

If H_o is true, then the distribution of

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

is a mixture of χ_1^2 and χ_0^2 distributions where we give equal weights to each (i.e., $1/2$).

Mixture of χ^2_0 & χ^2_1 with Equal Weights



Example of Case 1: HSB (using ML)

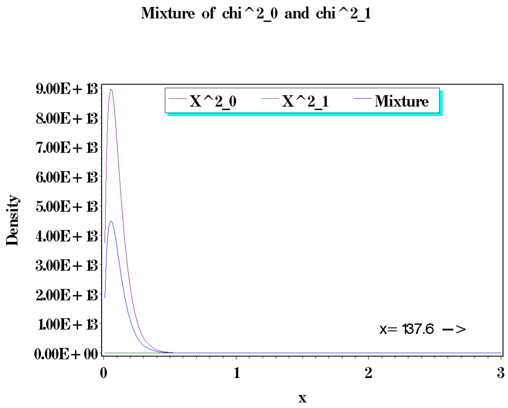
Null model: no random effects

Full model: random intercept.

		ML				
	No.	Deviance	Test		p-value from	
Model	of τ's	$-2\ln(\lambda_N)$	statistic	χ^2_0	χ^2_1	mixture
Null	0	46,372.3	137.6	0	$.89\text{E}-31$	$.45\text{E}-31$
Full	1	46,234.7	—			

The mixture p-value $= .5(.89\text{E}-31) = .45\text{E}-31.$

Mixture of χ^2_0 & χ^2_1 with Equal Weights



Case 2: One vs Two Random Effects

$$H_o : \boldsymbol{T} = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{versus} \quad H_a : \boldsymbol{T} = \begin{pmatrix} \tau_0^2 & \tau_{10} \\ \tau_{10} & \tau_1^2 \end{pmatrix}$$

In other words, we're testing

$$H_o : \tau_{10} = \tau_1^2 = 0 \quad \text{versus} \quad H_a : \text{not} \quad \tau_{10} = \tau_1^2 = 0$$

Assuming

- $\tau_0^2 > 0$ in H_o
- In H_a, $\boldsymbol{T}$ 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 χ_1^2 and χ_2^2 distributions.

Case 2 (continued)

To get the correct p-value we take a mixture of χ^2_1 and χ^2_2 distributions.

Case 2: HSB Example

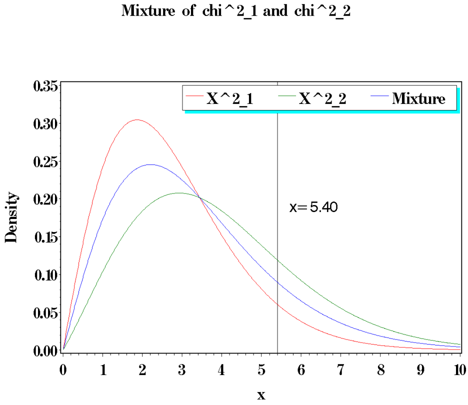
Null model: Random intercept only

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

		Maximum Likelihood				
	No.	Deviance	Test	p-value from		
Model	of τ's	$-2\ln(\lambda_N)$	statistic	χ^2_1	χ^2_2	mixture
Null	1	46,234.68	5.40	.020	.067	.04
Full	3	46,229.28	—			

Mixture p-value $= .5(.020) + .5(.067) = .04.$
Note Wald $p = .09.$

Mixture of χ^2_1 & χ^2_2 with Equal Weights



Case 3: q vs $q+1$ Random Effects

The hypotheses are

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

and

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

Case 3 (continued)

Assuming that

- In H_o, the $(q \times q)$ matrix of τ'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 χ^2_q and χ^2_{q+1}.

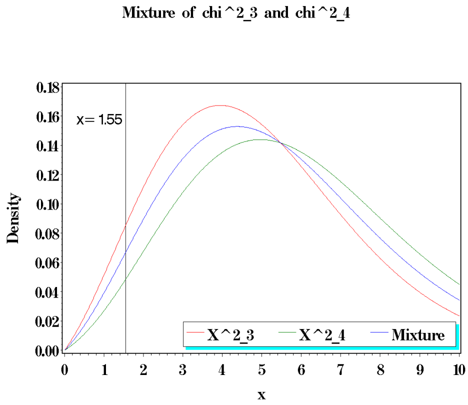
Case 3: HSB Example

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

Model	No. of τ's	Maximum Likelihood Deviance $-2\ln(\lambda_N)$	Test statistic	p-value from χ^2_3	χ^2_4	mixture
Null	6	46,225.12	1.55	.67	.82	.74
Full	10	46,223.56	—			

Mixture p-value $= .5(.67) + .5(.82) = .74$.

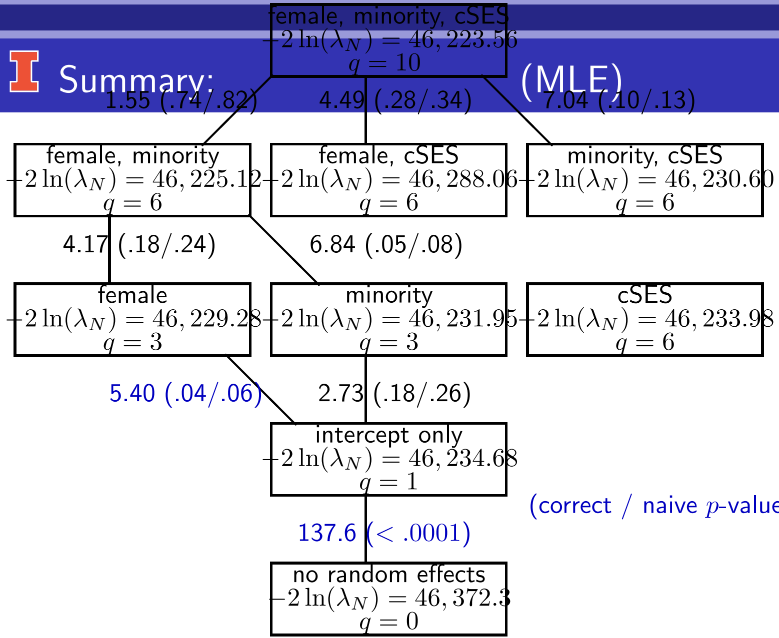
Mixture of χ^2_3 & χ^2_4 with Equal Weights



Case 4: q vs $q+k$ Random Effects

- The sample distribution of $-2\ln(\lambda_N)$ is a mixture of χ^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.

Summary: (MLE)



(correct / naive p-value)

Summary of The General Procedure

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

$$H_o : \tau^2_{q+1} = \tau_{q,q+1} = ... = \tau_{0,q+1} = 0 \text{ vs } H_a : \text{ not } H_0$$

- $\boldsymbol{T}$ must be a proper covariance matrix (i.e., $\tau^2_k > 0$ and $\tau_{kk'} \geq \tau_k\tau_{k'}$).
- Fit nested and full model.
- Compute likelihood ratio test statistic.
- Compare test statistic to χ^2_q and χ^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, $\boldsymbol{T}$ 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 $\hat{\boldsymbol{T}}$ is a "proper" covariance matrix (i.e., no correlations ≥ 1 or ≤ -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 $\boldsymbol{T}$ 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.

- Fvie 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 γ's, τ's and σ^2.
- N is the sample size.

Four Information Criteria

Criteria	Smaller-is-better	Reference
AIC	$-2\mathcal{L} + 2d$	Akaike (1974)
AICC	$-2\mathcal{L} + 2dn^*(n^* - d - 1)$	Hurvich & Tsai (1989)
		Burnham & Anderson (1998)
HQIC	$-2\mathcal{L} + 2d \log \log N$	Hannan & Quinn (1979)
BIC	$-2\mathcal{L} + d \log N$	Schwarz (1978)
CAIC	$-2\mathcal{L} + d \log(N + 1)$	Bozdogan (1987)

What is N?

- Number of groups/clusters $\rightarrow$ SAS
- Total number of observations $\rightarrow$ R lmer

What Should N be?

Delattre, M., Lavielle, M., Poursat, M.A. (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics*, *8, 456–475. DOI: 10.1214/140EJS890.*

Problem is that we have 2 levels and so neither the number of clusters nor total number of observations is ideal.

Starting from first principles, Delattre et al propose

$$BIC \approx -2\mathcal{L} + d_{random} \log(N) + d_{fixed} \log(n_{++})$$

where

- d_{random} *is number of variance and covariance parameters*
- d_{fixed} *is number of fixed effects parameters*
- N *number of clusters*
- n_{++} *total number of observations*

bic_hlm R function

This will compute AIC and 4 different versions of BIC

- bic.new ← deviance + nrandom*log(N) + nfixed*log(n_{++})
- bic.harm ← deviance + nrandom*log(N) + nfixed*log(N$\bar{n_j}$) where $\bar{n_j}$ is the harmonic mean
- bic.ngrps ← deviance + nrandom*log(N) (i.e., SAS)
- bic.ntot ← deviance + nrandom*log(n_{++}) (i.e., lmer)

Use, for example, the hsb data set

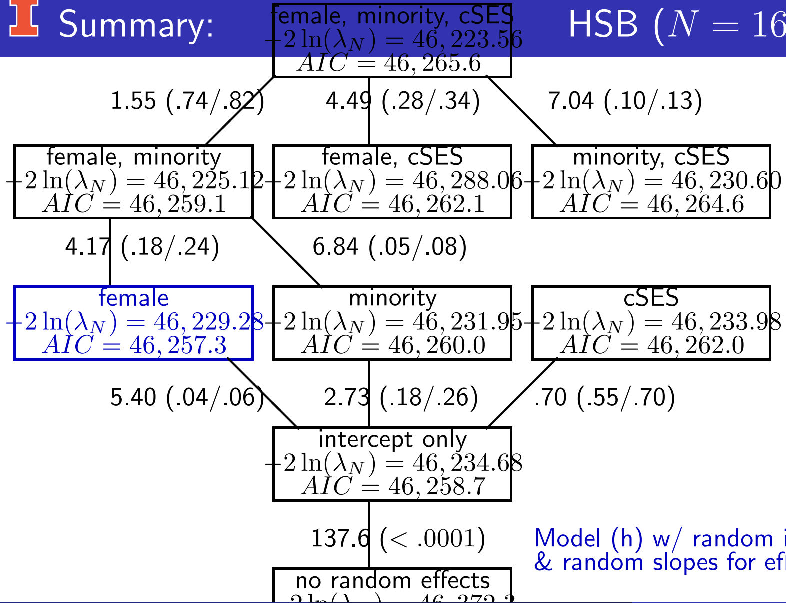
```
bic.hlm(model1,hsb$id)
```

	AIC	bic.new	bic.harm	bic.ngrps	bic.ntot
Model 1	46307.34	46361.65	46361.02	46335.02	46369.26
Model 2	46301.58	46362.76	46362.05	46332.33	46370.37
Model 3	46268.70	46384.20	46382.85	46327.13	46399.42

Notes Regarding Information Criteria

- Be sure that you know whether the one you're using is a "larger" or "smaller–is–better."
- SAS/MIXED and R give "smaller–is– better"
- Information criteria are only "rules of thumb" and not statistical tests.
  - Difference 0-1, not important difference
  - 2-3 moderate difference
  - ≥ 4 "big" difference
- The different criteria many not always agree.

Summary: HSB ($N = 160$)



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.

R^2 type measures

In multiple regression, there are different of 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.

R^2 Measures in Multilevel Models:

We need to consider micro and macro level residual variance.

So we need to propose measures for each level:

- R_1^2: level 1
- R_2^2: level 2

Proportional Reduction Unexplained Variance

$$\text{For level 2: } R_2^2 = \frac{\tau_0^{*2}}{\tau_0^2}$$

where

- τ_0^{*2} is level 2 residual variance with predictor variables (micro and/or macro) in the model.
- τ_0^2 is without predictor variables.

This value can be greater than one; that is, when $\tau_0^{*2} > \tau_0^2$, $R_2^2 => 1$. (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

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

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},

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

where

- The γ's are fixed (considered to be known).
- The X_k's are random variables with means μ_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 $X_{k,ij} = x_{k,ij}$.

$$\begin{aligned} \mathsf{E}(Y_{ij}|X_{k,ij} = x_{k,ij}) &= \mathsf{E}\left[\sum_{k=0}^{p} \gamma_{k0} x_{k,ij} + U_{0j} + R_{ij}\right] \\ &= \sum_{k=0}^{p} \gamma_{k0} x_{k,ij} + \mathsf{E}[U_{0j}] + \mathsf{E}[R_{ij}] \\ &= \sum_{k=0}^{p} \gamma_{k0} x_{k,ij} \end{aligned}$$

. . . and get $\hat{\tau}_0^2$ and $\hat{\sigma}^2$.

Level 1, Case 2: Estimation

The estimated mean squared error of prediction,

$$\begin{aligned}
\frac{1}{n_+}\sum_{j=1}^{N}\sum_{i=1}^{n_j}(Y_{ij}-\hat{Y}_{ij})^2 &= \text{var}(Y_{ij}-\hat{Y}_{ij}) \\
&= \text{var}(Y_{ij}-\sum_{k=0}^{p}\hat{\gamma}_{k0}x_{k,ij}) \\
&= \hat{\tau}_0^2+\hat{\sigma}^2
\end{aligned}$$

The Level 1 Measure

$$
\begin{aligned}
R_1^2 &= \frac{\mathsf{var}(Y_{ij}) - \mathsf{var}(Y_{ij} - \sum_{k=0}^{p} \gamma_{k0} x_{k,ij})}{\mathsf{var}(Y_{ij})} \\
&= 1 - \frac{\mathsf{var}(Y_{ij} - \sum_{k=0}^{p} \gamma_{k0} x_{k,ij})}{\mathsf{var}(Y_{ij})} \\
&= 1 - \frac{\tau_0^2 + \sigma^2}{\tau_0^{*2} + \sigma^{*2}}
\end{aligned}
$$

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

Model		$\hat{\tau}_0^2$	$\hat{\sigma}^2$	$\hat{\tau}_0^2 + \hat{\sigma}^2$	R_1^2
(a)	null	8.55	39.15	47.70	
(b)	cSES	8.61	37.01	45.61	.044
(c)	(b) + minority	6.64	36.12	42.77	.103
(d)	(c) + female	6.26	35.88	42.14	.117
(e)	(d) + sector, size,				
	& meanSES	1.61	35.89	37.50	.214
(h)	(e) + cSES∗sector				
	female∗size,				
	& minority∗sector	1.67	35.59	37.26	.219
(i)	(h) + 5 more cross-level	1.63	35.59	37.22	.220

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

$$R_2^2 = 1 - \frac{\sigma^2/n + \tau_0^2}{\sigma^{*2}/n + \tau_0^{*2}} \qquad \frac{\longleftarrow \text{ with predictors}}{\longleftarrow \text{ null}}$$

R_2^2 Type Measure

$$
\begin{aligned}
R_2^2 &= \frac{\text{var}(\bar{Y}_{+j}) - \text{var}(\bar{Y}_{+j} - \sum_{k=1}^{p}\gamma_{k0}\bar{x}_{k,+j})}{\text{var}(\bar{Y}_{+j})} \\
&= 1 - \frac{\text{var}(\bar{Y}_{+j} - \sum_{k=1}^{p}\gamma_{k0}\bar{x}_{k,+j})}{\text{var}(\bar{Y}_{+j})} \\
&= 1 - \frac{\sigma^2/n + \tau_0^2}{\sigma^{*2}/n + \tau_0^{*2}}
\end{aligned}
$$

where

- σ^2 and τ_0^2 are from the model with predictor variables.
- σ^{*2} and τ^{*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 values (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 number of macro units.
- n_j is the number of cases within macro unit j.

HSB example

The harmonic mean equals 41.0587.

Model		$\hat{\tau}_0^2$	$\hat{\sigma}^2$	R_1^2	R_2^2
(a)	null	8.555	39.149		
(b)	cSES	8.612	37.005	.045	.006
(c)	(b) + minority	6.648	36.121	.105	.213
(d)	(c) + female	6.264	35.877	.118	.254
(e)	(d) + sector, size, meanSES	1.610	35.890	.215	.740
(h)	(e) + cSES$*$sector, female$*$size minority$*$sector	1.668	35.594	.220	.735
(i)	(h) + five more cross-level	1.631	35.589	.221	.739

R^2s 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.
- Recchia, A. (2010). R-Squared measures for two-level hierarchial linear models Using SAS. *JSS, 31*, Code Snippet 2. URL: http://www.jstatsoft.org/v32/c02/paper.
- If you use this MACRO, use the reference above and the date when the macro was downloaded.
  For lmer, I wrote a function to do this called hlmRsq, which is included in the file
  "All$_{,.01in}functions.txt''thatisonecourseweb - site.$

HSB example

Harmonic mean: $n = 41.0587$

Model	$\hat{\tau}_0^2$	$\hat{\sigma}^2$	$\hat{\tau}_0^2 + \hat{\sigma}^2$	R_1^2	R_2^2
Random Intercept Models					
null	8.55	39.15	47.70		
$(\text{SES}_{ij} - \overline{\text{SES}}_j)$	8.61	37.01	45.62	.05	.01
$+\overline{\text{SES}}_j$	2.65	37.01	39.66	.17	.63
$+\text{Female}_{ij} + \text{Minority}_{ij}$					
$+\text{pAcademic}_j + \text{Sector}_j$	1.53	35.89	37.04	.22	.75
Random Slope/Effects Models		$(\bar{n}_+ = 41.06)$			
$(\text{SES}_{ij} - \overline{\text{SES}}_j) + \overline{\text{SES}}_j$				.17	.63
$+\text{Female}_{ij} + \text{Minority}_{ij}$					
$+\text{pAcademic}_j + \text{Sector}_j$				.22	.75
$+\text{Minority}_{ij}$ (random					
intercept dropped)				.22	.74

SAS: 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:

```
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:

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

5 Output:

Explained Proportion of Variance

Rep Size	Level 1	Level 2
41.06	0.212868	0.733890

lmer and R^2s

- I could not find a function or package that does this so I wrote a function called `hlmRsq`.
- Download it from the course website (it's in file "All$_{,.01in}Functions.txt''.Use$"source"$tellRwherethefileisandit'snam$
- Fit a model: `model1 ← lmer(mathach ∼ 1 + cses + meanSES + (1 + cses | id ), data=hsb, REML=FALSE)`
  IF YOU HAVE RANDOM SLOPES, PUT THEM IN FIXED PART OF MODEL FIRST.
- Run:

```
hlmRsq(hsb, model1, hsb$Id)
   harmonic.mean       R1sq       R2sq
        41.05874  0.2191968  0.7387011
```

Notes: R^2 Measures for HLMs

- In the population, R_1^2 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_1^2 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, $\mathsf{E}(\boldsymbol{U}_j) = \mathbf{0}$, and X_{ij}, $\boldsymbol{U}_j$ and R_{ij} are independent of each other.

One Final Comment: Like the information criteria, R_1^2 and R_2^2 are indices of fit and are not statistical significance tests.

Summary

- Tests for fixed effects:
  - Wald, t and F tests $\longrightarrow$ OK under MLE and REML.
  - Likelihood ratio only valid under MLE.
- Test for random effects:
  - Testing $H_o : \tau^2 = 0$ 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 χ^2_{q+1} and χ^2_q.
- Global measures:
  - Information criteria: useful for model comparison.
  - R_1^2 and R_2^2: can detect model miss-specification.