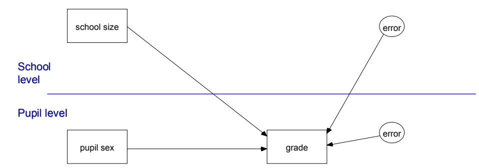
Sample Size & Robustness Issues in Multilevel Regression Analysis

Joop Hox & Cora Maas
Utrecht University, the Netherlands

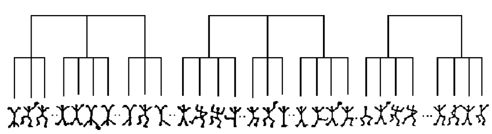
j.hox@fss.uu.nl
http://www.fss.uu.nl/ms/jh

Graphical Picture of Simple Two-level Regression Model



- Outcome variable on pupil level
- Explanatory variables at both levels
- Residual error at individual level
- *Residual error at school level*

Hierarchical Data Structure



- Three level data structure
- Groups at different levels may have different sizes
- Response (outcome) variable at lowest level
- Explanatory variables at all levels
- The statistical model assumes *sampling* at all levels

Multilevel Regression Model

- Explanatory variables at all levels
- Higher level variables predict variation of lowest level intercept and slopes
    - At the lowest (individual) level we have
    - $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij}$
    - and at the second level
    - $\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j}$ and $\beta_{1j} = \gamma_{10} + \gamma_{11} Zj + u_{1j}$
- Hence

  $Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$

Estimation

- Multilevel regression model:

  $Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$

- Maximum Likelihood (ML) estimation
    - Gamma coefficients
        - standard errors, *p* -values
    - Variance components s_e^2 and s^2_{u0}, s_{u01}, s^2_{u1}
        - standard errors, *p* -values
- Question: how accurate is the ML estimation with small samples and/or lack of normality?

Sample size and Accuracy

- Fixed parameters (regression coefficients)
    - Unbiased
    - Standard errors generally accurate
- Random parameters (variance components)
    - Lowest level: fine
    - Higher levels: problems with small samples
- Simulations suggest many groups more important than many individuals
- Simulations largely unpublished; reviewed by Kreft (1996) and Hox (1998), summarized in Hox (2002)

Two Questions

- What are sufficient sample sizes for multilevel regression modeling
  - Sufficient = accurate estimates and standard errors
- What happens when higher-level errors are not normal?
  - At what sample sizes is ML robust?
    - robust = accurate estimates and standard errors
  - Are robust standard errors indeed better?
- Answers by simulation

Simulation Model

- Simple simulation model
  - $Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$
- Note: in null model

  $Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$

  we can estimate variances σ^2_{u0} and σ^2_e giving intraclass correlation (ICC)

  $\rho_I = \sigma^2_{u0} / (\sigma^2_{u0} + \sigma^2_e)$
- Previous simulations show ICC important

Simulation Design

- Simple simulation model
  - $Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$
- Simulated conditions:

1) NG = Number of Groups [30, 50, 100]
2) GS = Group Size [5, 30, 50]
3) ICC = Intraclass Correlation [.1, .2, .3]

- 27 simulated conditions, 1000 simulated data sets in each condition, residuals normal distribution

Simulation Results

- *No* convergence problems: 27000 admissible solutions
- Regression coefficients *no* detectable bias
- Variances almost no bias
  - In condition NG=30, GS=5, ICC=0.3 bias = 0.3%
- Standard errors
  - *No* effect of ICC
  - For regression coefficients: accurate
  - For variances:

    Small effect of Number of Groups and Group Size

Simulation Results: Effect of Number of Groups

Parameter	Number of Groups			
	30	50	100	*p*-value
U_0	0.089	0.074	0.060	.0000
U_1	0.088	0.072	0.057	.0000
E_0	0.058	0.056	0.049	.0102

- If NG = 100 only U_0 and U_1 coverage significantly different from 95%

Simulation Results: Effect of Group Size

Influence of the Group Size on the non-coverage of the 95% confidence interval

Parameter	Group Size			
	5	30	50	*p*-value
U_0	0.074	0.075	0.074	.9419
U_1	0.078	0.066	0.072	.0080
E_0	0.061	0.051	0.051	.0055

- If GS = 100 only U_0 and U_1 coverage significantly different from 95%

Simulation Results: Conclusions

- Point estimates of regression coefficients *fine*
- Point estimates of variance components *fine*

- Standard errors for regression coefficients *fine*

- Standard errors for variance components show bias when Number of Groups < 100
  - If NG = 50, bias 8%
  - If NG = 30, bias 15%

13

Simulation Design

- Simple simulation model
  - $Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} Z_j X_{ij} + u_{1j} X_{ij} + u_{0j} + e_{ij}$

- Simulated conditions:

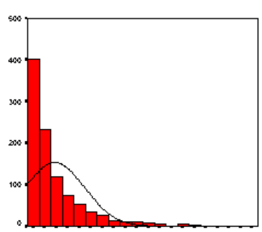
1) NG = Number of Groups [30, 50, 100]
2) GS = Group Size [5, 30, 50]
3) ICC = Intraclass Correlation [.1, .2, .3]

- 1000 simulated data sets in each condition, but residuals U_0 and U_1 have χ_1^2 distribution

14

Simulation Design

- 1000 simulated data sets in each condition, but residuals U_0 and U_1 have χ_1^2 distribution



15

Robust Standard Errors

- At what sample sizes is ML robust?
- Are robust standard errors indeed better?

- Robust standard errors (MLwiN, AML): sandwich estimates or Huber/White estimates

- ML sampling variance: $\mathbf{V}_A(\hat{\beta}) = \mathbf{H}^{-1}$

- Sandwich sampling variance: $\mathbf{V}_R(\hat{\beta}) = \mathbf{H}^{-1}\mathbf{C}\mathbf{H}^{-1}$
  C based on observed residuals

16

Simulation Results

- *No* convergence problems: 27000 admissible solutions

- Regression coefficients *no* detectable bias
- Variances almost no bias
  - In condition NG=30, GS=5, ICC=0.1 bias = -0.1%

- Standard errors
  - No effect of ICC
  - For regression coefficients: small negative bias, robust s.e. not better
  - For variances: some large negative biases, robust s.e. better, but not good enough

17

Simulation Results: Effect of Number of Groups

- Actual coverage of nominal 95% C.I.: ML/SW

	E_0	U_0	U_1
NG 30	.9487/.9866*	.6537*/.8128*	.6501*/.8007*
50	.9539/.9903*	.6701*/.8734*	.6471*/.8506*
100	.9534/.9933*	.6659*/.9217*	.6308*/.9059*

- Huge negative bias for 2nd level variance components: standard errors estimated much too small
- Robust standard errors better, but not good enough
  - With 200 groups probably good enough

18

Simulation Results: Effect of Group Size

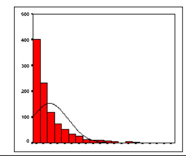
- Actual coverage of nominal 95% C.I.: ML/SW

		E_0	U_0	U_1
GS	5	.9373/.9819*	.7784*/.9019*	.7540*/.8648*
	30	.9630/.9937*	.6219*/.8582*	.6032*/.8500*
	50	.9557/.9947*	.5893*/.8478*	.5708*/.8423*

- Huge negative bias for 2nd level variance components: standard errors estimated much too small
- Robust standard errors better, but not good enough
- Having larger groups makes problem *worse*

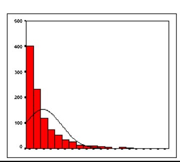
Conclusions: Regression Coefficients

- Point estimates fine in all simulated conditions
- Standard errors fine in all simulated conditions with normality assumption valid
- Standard errors too small with normality assumption violated
  - With NG=100 coverage good
  - Robust standard errors not better
- Advice: don't worry, be happy



Conclusions: Variance Components

- Point estimates fine in all simulated conditions
- Standard errors reasonable in all simulated conditions with normality assumption valid
- Standard errors much too small with normality assumption violated
  - Robust standard errors definitively better
  - But at NG = 100 not good enough
- Advice: use robust standard errors as diagnostic, not remedy



Thank You

Copies of transparencies on http://www.fss.uu.nl/ms/jh