MULTILEVEL ANALYSIS

Tom A. B. Snijders

http://www.stats.ox.ac.uk/~snijders/mlbook.htm

Department of Statistics
University of Oxford
2012

This is a set of slides following Snijders & Bosker (2012).

The page headings give the chapter numbers and the page numbers in the book.

Literature:

Tom Snijders & Roel Bosker,
Multilevel Analysis: An Introduction to Basic and Applied Multilevel Analysis,
2**nd** edition. Sage, 2012.
Chapters 1-2, 4-6, 8, 10, 13, 14, 17.

There is an associated website
http://www.stats.ox.ac.uk/~snijders/mlbook.htm
containing data sets and scripts for various software packages.

These slides are *not* self-contained, for understanding them it is necessary also to study the corresponding parts of the book!

2. Multilevel data and multilevel analysis

Multilevel Analysis using the hierarchical linear model :
random coefficient regression analysis for data with several nested levels.

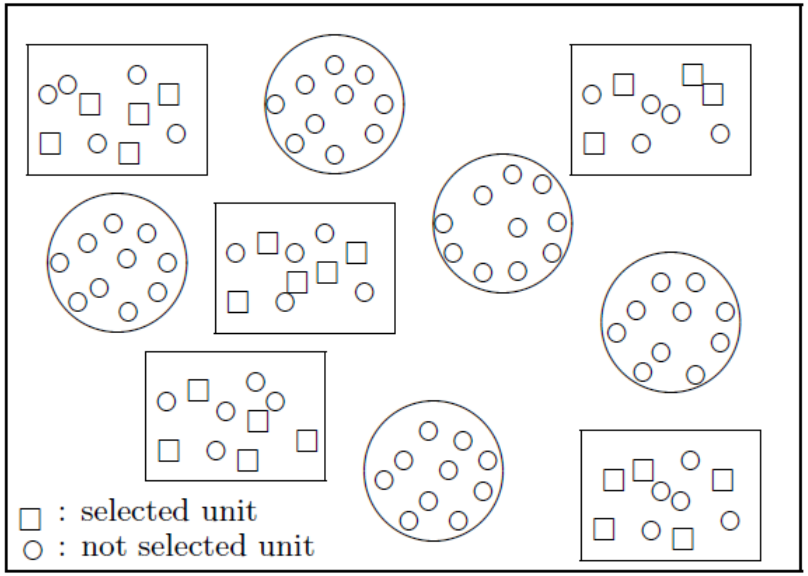


Figure 2.1: Multi-stage sample.

Each level is (potentially) a *source of unexplained variability*.

Some examples of units at the macro and micro level:

macro-level	micro-level
schools	teachers
classes	pupils
neighborhoods	families
districts	voters
firms	departments
departments	employees
families	children
litters	animals
doctors	patients
interviewers	respondents
judges	suspects
subjects	measurements
respondents = egos	alters

Multilevel analysis is a suitable approach to take into account the *social contexts* as well as the *individual respondents* or *subjects*.

The hierarchical linear model is a type of regression analysis for multilevel data where the dependent variable is at the lowest level.

Explanatory variables can be defined at any level (including aggregates of level-one variables).

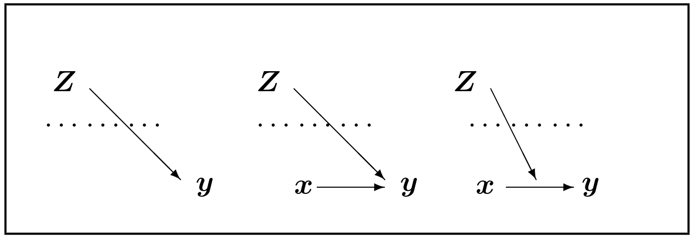


Figure 2.5 The structure of macro–micro propositions.

Also longitudinal data can be regarded as a nested structure; for such data the hierarchical linear model is likewise convenient.

Two kinds of argument to choose for a multilevel analysis instead of an OLS regression of disaggregated data:

1. *Dependence as a nuisance*
   Standard errors and tests base on OLS regression are suspect
   because the assumption of independent residuals is invalid.

2. *Dependence as an interesting phenomenon*
   It is interesting in itself to disentangle variability at the various levels;
   moreover, this can give insight in the directions
   where further explanation may fruitfully be sought.

4. The random intercept model

Hierarchical Linear Model:

$\boldsymbol{i}$ indicates level-one unit (e.g., individual);
$\boldsymbol{j}$ indicates level-two unit (e.g., group).

Variables for individual $\boldsymbol{i}$ in group $\boldsymbol{j}$:

$\boldsymbol{Y_{ij}}$ dependent variable;
$\boldsymbol{x_{ij}}$ explanatory variable at level one;

for group $\boldsymbol{j}$:

$\boldsymbol{z_j}$ explanatory variable at level two; $\boldsymbol{n_j}$ group size.

OLS regression model of $\boldsymbol{Y}$ on $\boldsymbol{X}$ ignoring groups :

$$\boldsymbol{Y_{ij} = \beta_0 + \beta_1 x_{ij} + R_{ij}} \ .$$

Group-dependent regressions:

$$\boldsymbol{Y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + R_{ij}} \ .$$

Distinguish two kinds of *fixed effects* models:

1. models where group structure is ignored;
2. models with fixed effects for groups: $\boldsymbol{\beta_{0j}}$ are fixed parameters.

In the *random intercept* model, the intercepts $\boldsymbol{\beta_{0j}}$ are random variables representing random differences between groups:

$$\boldsymbol{Y_{ij} = \beta_{0j} + \beta_1 x_{ij} + R_{ij}} \; .$$

where $\boldsymbol{\beta_{0j}}$ = average intercept γ_{00} plus group-dependent deviation $\boldsymbol{U_{0j}}$:

$$\boldsymbol{\beta_{0j}} = \gamma_{00} + \boldsymbol{U_{0j}} \; .$$

In this model, the regression coefficient $\boldsymbol{\beta_1}$ is common to all the groups.

In the random intercept model, the constant regression coefficient β_1 is sometimes denoted γ_{10}:

Substitution yields

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

In the hierarchical linear model, the U_{0j} are *random* variables.
The model assumption is that they are independent,
normally distributed with expected value 0, and variance

$$\tau^2 = \mathrm{var}(U_{0j}).$$

The statistical parameter in the model is not their individual values,
but their variance τ_0^2 .

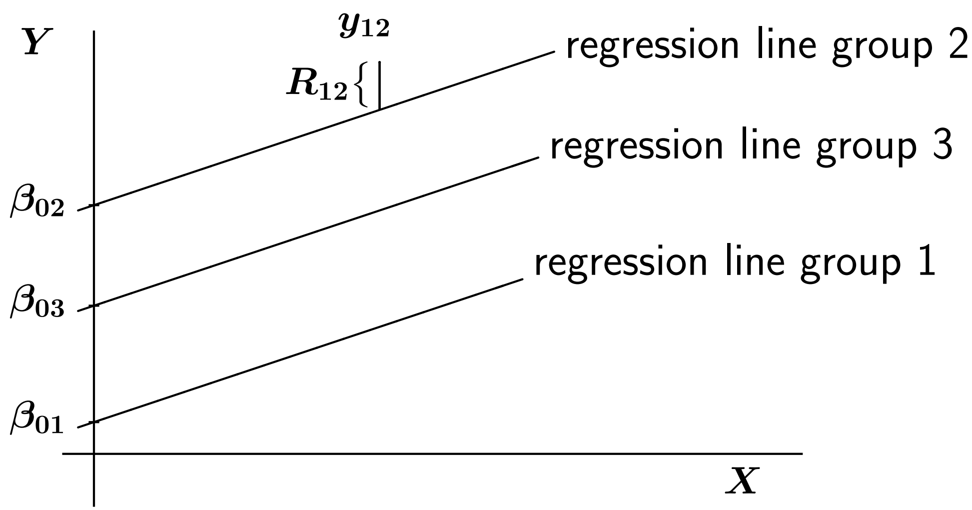


Figure 4.1 Different parallel regression lines.

The point $\boldsymbol{y_{12}}$ is indicated with its residual $\boldsymbol{R_{12}}$.

Arguments for choosing between fixed (F) and random (R) coefficient models for the group dummies:

1. If groups are unique entities and inference should focus on *these* groups: F .
   This often is the case with a small number of groups.
2. If groups are regarded as sample from a (perhaps hypothetical) population and inference should focus on this population, then R .
   This often is the case with a large number of groups.
3. If level-two effects are to be tested, then R .
4. If group sizes are small and there are many groups, and it is reasonable to assume exchangeability of group-level residuals, then R makes better use of the data.
5. If the researcher is interested only in *within-group* effects, and is suspicious about the model for *between-group* differences, then F is more robust.
6. If group effects $\boldsymbol{U_{0j}}$ (etc.) are not nearly normally distributed, R is risky
   (or use more complicated multilevel models).

The empty model (*random effects ANOVA*) is a model without explanatory variables:

$$\boldsymbol{Y}_{ij} = \gamma_{00} + \boldsymbol{U}_{0j} + \boldsymbol{R}_{ij} \; .$$

Variance decomposition:

$$\text{var}(\boldsymbol{Y}_{ij}) = \text{var}(\boldsymbol{U}_{0j}) + \text{var}(\boldsymbol{R}_{ij}) = \boldsymbol{\tau}_0^2 + \boldsymbol{\sigma}^2 \; .$$

Covariance between two individuals ($\boldsymbol{i} \neq \boldsymbol{i}'$) in the same group $\boldsymbol{j}$:

$$\text{cov}(\boldsymbol{Y}_{ij}, \boldsymbol{Y}_{i'j}) = \text{var}(\boldsymbol{U}_{0j}) = \boldsymbol{\tau}_0^2 \; ,$$

and their correlation:

$$\boldsymbol{\rho}(\boldsymbol{Y}_{ij}, \boldsymbol{Y}_{i'j}) = \boldsymbol{\rho}_{\boldsymbol{I}}(\boldsymbol{Y}) = \frac{\boldsymbol{\tau}_0^2}{(\boldsymbol{\tau}_0^2 + \boldsymbol{\sigma}^2)} \; .$$

This is the *intraclass correlation coefficient*.

Often between .05 and .25 in social science research, where the groups represent some kind of social grouping.

Example: 3758 pupils in 211 schools , Y = language test.

Classrooms / schools are level-2 units.

Table 4.1 Estimates for empty model

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.00	0.32

Random Part	Variance Component	S.E.
Level-two variance:		
$\tau_0^2 = \text{var}(U_{0j})$	18.12	2.16
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	62.85	1.49
Deviance	26595.3	

Intraclass correlation

$$\rho_I = \frac{18.12}{18.12 + 62.85} = 0.22$$

Total population of individual values Y_{ij} has estimated mean 41.00 and standard deviation $\sqrt{18.12 + 62.85} = 9.00$.

Population of class means β_{0j} has estimated mean 41.00 and standard deviation $\sqrt{18.12} = 4.3$.

The model becomes more interesting,
when also *fixed effects* of explanatory variables are included:

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

(Note the difference between fixed effects of explanatory variables and fixed effects of group dummies!)

Table 4.2 Estimates for random intercept model with effect for IQ

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.06	0.24
γ_{10} = Coefficient of IQ	2.507	0.054

Random Part	Variance Component	S.E.
Level-two variance:		
τ_0^2 = var(U_{0j})	9.85	1.21
Level-one variance:		
σ^2 = var(R_{ij})	40.47	0.96
Deviance	24912.2	

There are two kinds of parameters:

1. fixed effects: regression coefficients γ (just like in OLS regression);
2. random effects: variance components σ^2 and τ_0^2 .

Table 4.3 Estimates for ordinary least squares regression

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.30	0.12
γ_{10} = Coefficient of IQ	2.651	0.056

Random Part	Variance Component	S.E.
Level-one variance:		
$\sigma^2 = \mathrm{var}(R_{ij})$	49.80	1.15
Deviance	25351.0	

Multilevel model has more structure (“dependence interesting”);

OLS has misleading standard error for intercept (“dependence nuisance”).

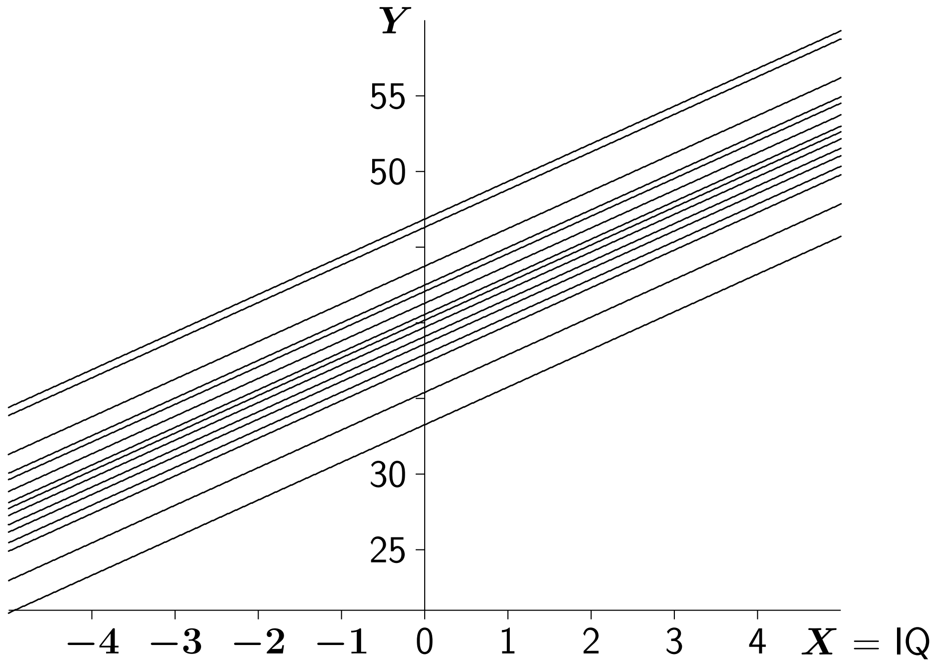


Figure 4.2 Fifteen randomly chosen regression lines according to the random intercept model of Table 4.2.

More explanatory variables:

$$\begin{aligned} \boldsymbol{Y}_{ij} = \gamma_{00} + \gamma_{10}\,\boldsymbol{x}_{1ij} + \ldots + \gamma_{p0}\,\boldsymbol{x}_{pij} + \gamma_{01}\,\boldsymbol{z}_{1j} + \ldots + \gamma_{0q}\,\boldsymbol{z}_{qj} \\ + \boldsymbol{U}_{0j} + \boldsymbol{R}_{ij} \;. \end{aligned}$$

Especially important:
difference between within-group and between-group regressions.

The within-group regression coefficient is the regression coefficient within each group, assumed to be the same across the groups.

The between-group regression coefficient is defined as the regression coefficient for the regression of the group means of $\boldsymbol{Y}$ on the group means of $\boldsymbol{X}$.

This distinction is essential to avoid *ecological fallacies* (p. 15–17 in the book).

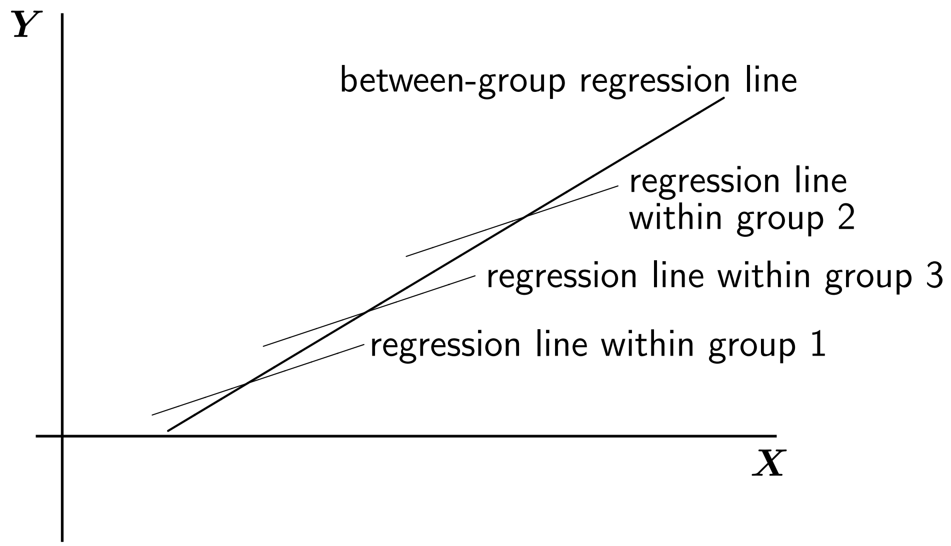


Figure 4.3 Different between-group and within-group regression lines.

This is obtained by having *separate fixed effects* for the level-1 variable $\boldsymbol{X}$ and its group mean $\bar{\boldsymbol{X}}$.
(Alternative:
use the within-group deviation variable $\tilde{\boldsymbol{X}}_{ij} = (\boldsymbol{X} - \bar{\boldsymbol{X}})$ instead of $\boldsymbol{X}$.)

Table 4.4 Estimates for random intercept model
with different within- and between-group regressions

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.11	0.23
γ_{10} = Coefficient of IQ	2.454	0.055
γ_{01} = Coefficient of $\overline{\text{IQ}}$ (group mean)	1.312	0.262

Random Part	Variance Component	S.E.
Level-two variance:		
$\tau_0^2 = \text{var}(U_{0j})$	8.68	1.10
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	40.43	0.96
Deviance	24888.0	

In the model with separate effects for the original variable x_{ij} and the group mean

$$Y_{ij} = \gamma_{00} + \gamma_{10}\, x_{ij} + \gamma_{01}\overline{x}_{.j} + U_{0j} + R_{ij} \ ,$$

the within-group regression coefficient is γ_{10} ,
between-group regression coefficient is $\gamma_{10} + \gamma_{01}$.
This is convenient because the difference between within-group and between-group coefficients can be tested by considering γ_{01}.

In the model with separate effects for group-centered variable $\tilde{x}_{ij}$
and the group mean

$$Y_{ij} = \tilde{\gamma}_{00} + \tilde{\gamma}_{10}\, \tilde{x}_{ij} + \tilde{\gamma}_{01}\overline{x}_{.j} + U_{0j} + R_{ij} \ ,$$

the within-group regression coefficient is $\tilde{\gamma}_{10}$,
the between-group regression coefficient is $\tilde{\gamma}_{01}$.
This is convenient because these coefficients are given immediately in the results, with their standard errors.

Both models are equivalent, and have the same fit: $\tilde{\gamma}_{10} = \gamma_{10}$, $\tilde{\gamma}_{01} = \gamma_{10} + \gamma_{01}$.

Estimation/prediction of random effects

The random effects U_{0j} are *not* statistical parameters and therefore they are not estimated as part of the estimation routine.

However, it sometimes is desirable to 'estimate' them. This can be done by the *empirical Bayes* method; these 'estimates' are also called the *posterior means*. In statistical terminology, this is not called 'estimation' but 'prediction', the name for the construction of likely values for unobserved random variables.

The posterior mean for group j is based on two kinds of information:

⇒ *sample information* : the data in group j;
⇒ *population information* : the value U_{0j} was drawn from a normal distribution with mean 0 and variance τ_0^2 .

The population information comes from the other groups.
If this information is reasonable, prediction is improved on average.

Suppose we wish to predict the 'true group mean' $\gamma_{00} + U_{0j}$.

The empirical Bayes estimate in the case of the empty model is a weighted average of the group mean and the overall mean:

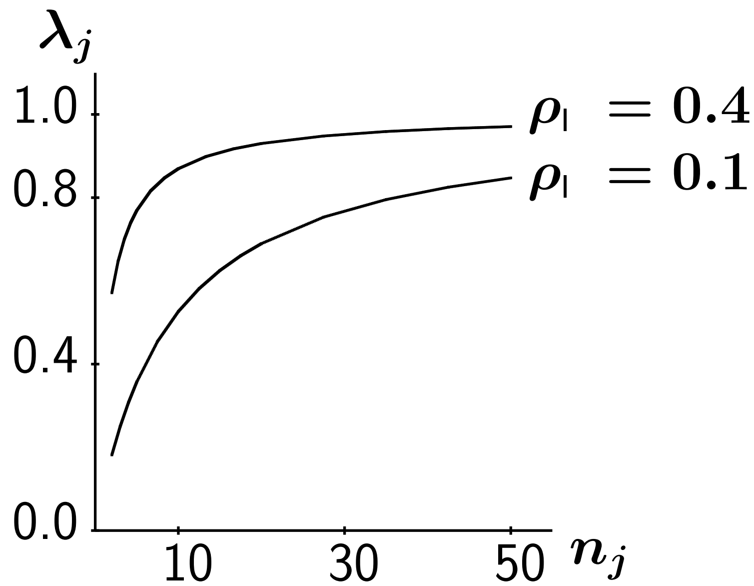
$$\hat{\beta}_{0j}^{\mathrm{EB}} = \lambda_j \, \bar{Y}_{.j} \; + \; (1 - \lambda_j) \, \hat{\gamma}_{00} \; ,$$

where the weight λ_j is the 'reliability' of the mean of group j

$$\lambda_j \; = \; \frac{\tau_0^2}{\tau_0^2 \; + \; \sigma^2 / n_j} \; = \; \frac{n_j \rho_{\mathrm{I}}}{1 \; + \; (n_j - 1) \rho_{\mathrm{I}}} \; .$$

The reliability coefficient indicates how well the true mean $\gamma_{00} + U_{0j}$ is measured by the observed mean $\bar{Y}_{.j}$; see Section 3.5.

The picture to the rights gives a plot.



These ‘estimates’ are not unbiased for each specific group, but they are more precise when the mean squared errors are averaged over all groups.

For models with explanatory variables, the same principle can be applied:
the values that would be obtained as OLS estimates per group are
“shrunk towards the mean”.

The empirical Bayes estimates, also called posterior means,
are also called shrinkage estimators.

There are two kinds of standard errors for empirical Bayes estimates:

comparative standard errors

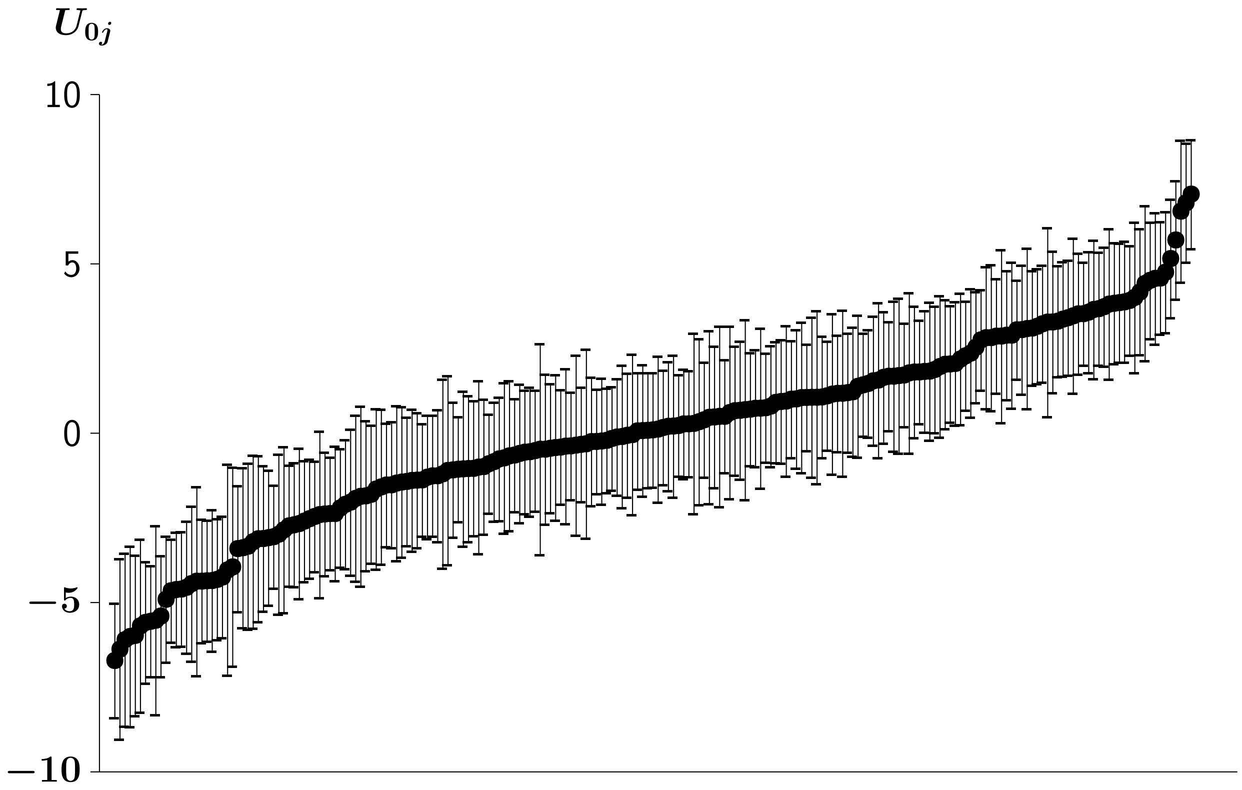
$$\text{S.E.}_{\text{comp}}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}}\right) = \text{S.E.}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}} - \boldsymbol{U}_{\boldsymbol{hj}}\right)$$

for comparing the random effects of different level-2 units
(use with caution – E.B. estimates are not unbiased!);

and *diagnostic standard errors*

$$\text{S.E.}_{\text{diag}}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}}\right) = \text{S.E.}\left(\hat{\boldsymbol{U}}^{\text{EB}}_{\boldsymbol{hj}}\right)$$

used for model checking (e.g., checking normality of the level-two residuals).



The ordered added value scores for 211 schools with comparative posterior confidence intervals.

In this figure, the error bars extend 1.39 times the comparative standard errors to either side, so that schools may be deemed to be significantly different if the intervals do not overlap (no correction for multiple testing!).

5. The hierarchical linear model

It is possible that not only the group average of $\boldsymbol{Y}$,
but also the effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is *randomly* dependent on the group.

In other words, in the equation

$$\boldsymbol{Y_{ij}} = \boldsymbol{\beta_{0j}} + \boldsymbol{\beta_{1j}}\, \boldsymbol{x_{ij}} + \boldsymbol{R_{ij}} \ ,$$

also the regression coefficient $\boldsymbol{\beta_{1j}}$ has a random part:

$$\begin{aligned} \boldsymbol{\beta_{0j}} &= \gamma_{00} + \boldsymbol{U_{0j}} \\ \boldsymbol{\beta_{1j}} &= \gamma_{10} + \boldsymbol{U_{1j}} \ . \end{aligned}$$

Substitution leads to

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

Variable $\boldsymbol{X}$ now has a *random slope*.

It is assumed that the group-dependent coefficients $(U_{0j},\ U_{1j})$ are independent across j, with a bivariate normal distribution with expected values $(0, 0)$ and covariance matrix defined by

$$\begin{aligned} \text{var}(U_{0j}) &= \tau_{00} = \tau_0^2\ ; \\ \text{var}(U_{1j}) &= \tau_{11} = \tau_1^2\ ; \\ \text{cov}(U_{0j}, U_{1j}) &= \tau_{01}\ . \end{aligned}$$

Again, the $(U_{0j},\ U_{1j})$ are not individual parameters in the statistical sense, but only their variances, and covariance, are the parameters.

Thus we have a linear model for the mean structure, and a parametrized covariance matrix within groups with independence between groups.

5.1 Estimates for random slope model

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.127	0.234
γ_{10} = Coeff. of IQ	2.480	0.064
γ_{01} = Coeff. of $\overline{\text{IQ}}$ (group mean)	1.029	0.262

Random Part	Parameters	S.E.
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.877	1.117
$\tau_1^2 = \text{var}(U_{1j})$	0.195	0.076
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	−0.835	0.217
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	39.685	0.964
Deviance	24864.9	

$\overline{\text{IQ}}$ is defined as the group mean.

The equation for this table is

$$Y_{ij} = 41.13 + 2.480\,\text{IQ}_{ij} + 1.029\,\overline{\text{IQ}}_{.j} + U_{0j} + U_{1j}\,\text{IQ}_{ij} + R_{ij}\,.$$

The slope β_{1j} has average 2.480 and s.d. $\sqrt{0.195} = 0.44$.

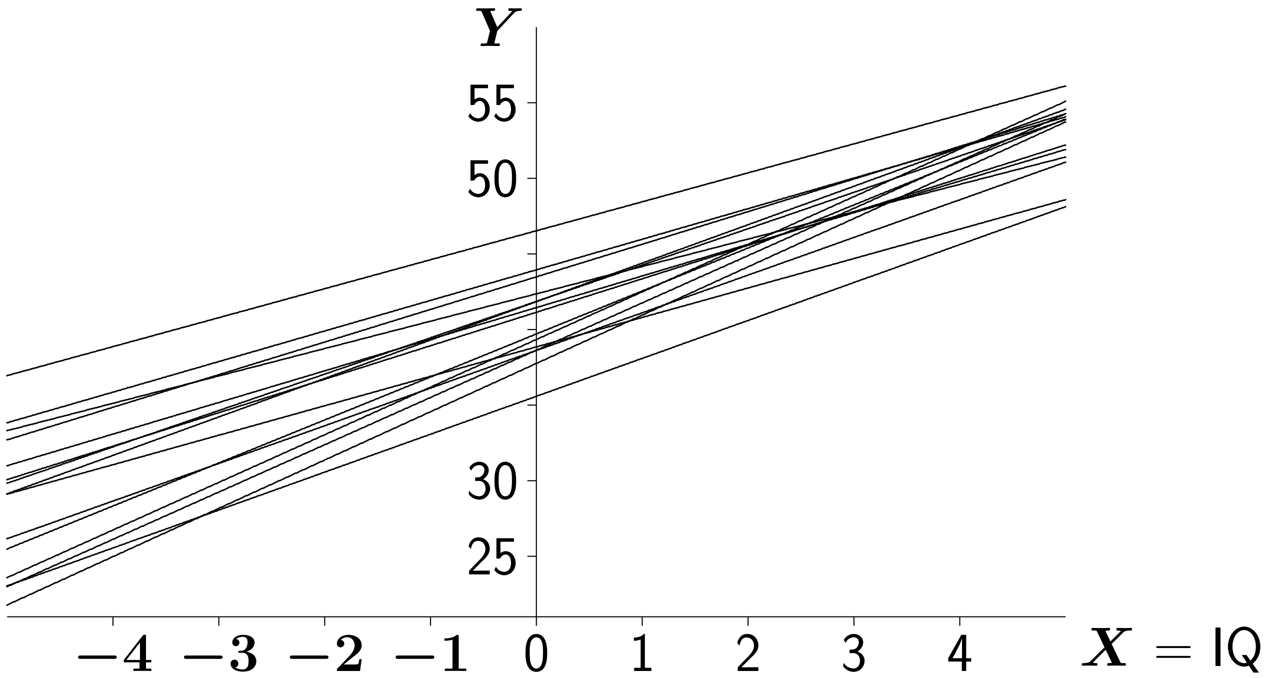


Figure 5.2 Fifteen random regression lines according to the model of Table 5.1.

Note the heteroscedasticity: variance is larger for low X than for high X.
The lines fan in towards the right.

Intercept variance and intercept-slope covariance depend on the position of the $X = 0$ value, because the intercept is defined by the $X = 0$ axis.

The next step is to *explain* the random slopes:

$$\boldsymbol{\beta}_{0j} = \gamma_{00} + \gamma_{01}\, \boldsymbol{z}_j + \boldsymbol{U}_{0j}$$

$$\boldsymbol{\beta}_{1j} = \gamma_{10} + \gamma_{11}\, \boldsymbol{z}_j + \boldsymbol{U}_{1j} \; .$$

Substitution then yields

$$\begin{aligned} \boldsymbol{Y}_{ij} &= (\gamma_{00} + \gamma_{01}\, \boldsymbol{z}_j + \boldsymbol{U}_{0j}) \\ &\qquad + (\gamma_{10} + \gamma_{11}\, \boldsymbol{z}_j + \boldsymbol{U}_{1j})\, \boldsymbol{x}_{ij} + \boldsymbol{R}_{ij} \\ &= \gamma_{00} + \gamma_{01}\, \boldsymbol{z}_j + \gamma_{10}\, \boldsymbol{x}_{ij} + \gamma_{11}\, \boldsymbol{z}_j\, \boldsymbol{x}_{ij} \\ &\qquad\qquad + \boldsymbol{U}_{0j} + \boldsymbol{U}_{1j}\, \boldsymbol{x}_{ij} + \boldsymbol{R}_{ij} \; . \end{aligned}$$

The term $\gamma_{11}\, \boldsymbol{z}_j\, \boldsymbol{x}_{ij}$ is called the *cross-level interaction effect*.

Table 5.2 Estimates for model with random slope and cross-level interaction

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.254	0.235
γ_{10} = Coefficient of IQ	2.463	0.063
γ_{01} = Coefficient of $\overline{\text{IQ}}$	1.131	0.262
γ_{11} = Coefficient of $\overline{\text{IQ}} \times$ IQ	–0.187	0.064

Random Part	Parameters	S.E.
Level-two random part:		
$\boldsymbol{\tau}_0^2 = \text{var}(\boldsymbol{U}_{0j})$	8.601	1.088
$\boldsymbol{\tau}_1^2 = \text{var}(\boldsymbol{U}_{1j})$	0.163	0.072
$\boldsymbol{\tau}_{01} = \text{cov}(\boldsymbol{U}_{0j}, \boldsymbol{U}_{1j})$	–0.833	0.210
Level-one variance:		
$\boldsymbol{\sigma}^2 = \text{var}(\boldsymbol{R}_{ij})$	39.758	0.965
Deviance	24856.8	

For two variables (IQ and SES) and two levels (student and school), the main effects and interactions give rise to a lot of possible combinations:

Table 5.3 Estimates for model with random slopes and many effects

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.632	0.255
γ_{10} = Coefficient of IQ	2.230	0.063
γ_{20} = Coefficient of SES	0.172	0.012
γ_{30} = Interaction of IQ and SES	–0.019	0.006
γ_{01} = Coefficient of $\overline{\text{IQ}}$	0.816	0.308
γ_{02} = Coefficient of $\overline{\text{SES}}$	–0.090	0.044
γ_{03} = Interaction of $\overline{\text{IQ}}$ and $\overline{\text{SES}}$	–0.134	0.037
γ_{11} = Interaction of IQ and $\overline{\text{IQ}}$	–0.081	0.081
γ_{12} = Interaction of IQ and $\overline{\text{SES}}$	0.004	0.013
γ_{21} = Interaction of SES and $\overline{\text{IQ}}$	0.023	0.018
γ_{22} = Interaction of SES and $\overline{\text{SES}}$	0.000	0.002

(continued next page....)

Random Part	Parameters	S.E.
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.344	1.407
$\tau_1^2 = \text{var}(U_{1j})$	0.165	0.069
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	–0.942	0.204
$\tau_2^2 = \text{var}(U_{2j})$	0.0	0.0
$\tau_{02} = \text{cov}(U_{0j}, U_{2j})$	0.0	0.0
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	37.358	0.907
Deviance	24624.0	

The non-significant parts of the model may be dropped:

Table 5.4 Estimates for a more parsimonious model with a random slope and many effects

Fixed Effect	Coefficient	S.E.
γ_{00} = Intercept	41.612	0.247
γ_{10} = Coefficient of IQ	2.231	0.063
γ_{20} = Coefficient of SES	0.174	0.012
γ_{30} = Interaction of IQ and SES	–0.017	0.005
γ_{01} = Coefficient of $\overline{\text{IQ}}$	0.760	0.296
γ_{02} = Coefficient of $\overline{\text{SES}}$	–0.089	0.042
γ_{03} = Interaction of $\overline{\text{IQ}}$ and $\overline{\text{SES}}$	–0.120	0.033
Random Part	**Parameters**	**S.E.**
Level-two random part:		
$\tau_0^2 = \text{var}(U_{0j})$	8.369	1.050
$\tau_1^2 = \text{var}(U_{1j})$	0.164	0.069
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	–0.929	0.204
Level-one variance:		
$\sigma^2 = \text{var}(R_{ij})$	37.378	0.907
Deviance	24626.8	

General formulation of the two-level model

As a link to the general statistical literature,
it may be noted that the two-level model can be expressed as follows:

$$\boldsymbol{Y}_j = \boldsymbol{X}_j\,\gamma + \boldsymbol{Z}_j\boldsymbol{U}_j + \boldsymbol{R}_j$$

with $\begin{bmatrix}\boldsymbol{R}_j\\ \boldsymbol{U}_j\end{bmatrix} \sim \boldsymbol{\mathcal{N}}\left(\begin{bmatrix}\emptyset\\ \emptyset\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_j(\boldsymbol{\theta}) & \emptyset\\ \emptyset & \boldsymbol{\Omega}(\boldsymbol{\xi})\end{bmatrix}\right)$

and $(\boldsymbol{R}_j, \boldsymbol{U}_j) \perp (\boldsymbol{R}_\ell, \boldsymbol{U}_\ell)$ for all $\boldsymbol{j} \neq \boldsymbol{\ell}$.

Standard specification $\boldsymbol{\Sigma}_j(\boldsymbol{\theta}) = \boldsymbol{\sigma}^2\boldsymbol{I}_{n_j}$,
but other specifications are possible.
Mostly, $\boldsymbol{\Sigma}_j(\boldsymbol{\theta})$ is diagonal, but even this is not necessary (e.g. time series).

The model formulation yields

$$\boldsymbol{Y}_j \sim \mathcal{N}\left(\boldsymbol{X}_j\gamma,\ \boldsymbol{Z}_j\boldsymbol{\Omega}(\boldsymbol{\xi})\boldsymbol{Z}_j' \ + \ \boldsymbol{\Sigma}_j(\boldsymbol{\theta})\right) \ .$$

This is a special case of the *mixed linear model*

$$\boldsymbol{Y} = \boldsymbol{X}\gamma + \boldsymbol{Z}\boldsymbol{U} + \boldsymbol{R},$$

with $\boldsymbol{X}[\boldsymbol{n},\boldsymbol{r}]$, $\boldsymbol{Z}[\boldsymbol{n},\boldsymbol{p}]$, and

$$\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{U}\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\emptyset\\ \emptyset\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma} & \emptyset\\ \emptyset & \boldsymbol{\Omega}\end{pmatrix}\right).$$

For estimation, the ML and REML methods are mostly used.

These can be implemented by various algorithms: Fisher scoring,
EM = Expectation–Maximization, IGLS = Iterative Generalized Least Squares.
See Section 4.7 and 5.4.

6. Testing

To test fixed effects, use the t-test with test statistic

$$T(\gamma_h) = \frac{\hat{\gamma}_h}{\text{S.E.}(\hat{\gamma}_h)} \; .$$

(Or the Wald test for testing several parameters simultaneously.)
The standard error should be based on REML estimation.

Degrees of freedom for the t-test, or the denominator of the F-test:
For a level-1 variable: $M - r - 1$,
where M = total number of level-1 units, r = number of tested level-1 variables.

For a level-2 variable: $N - q - 1$,
where N = number of level-2 units, q = number of tested level-2 variables.

For a cross-level interaction: again $N - q - 1$,
where now q = number of other level-2 variables interacting with this level-1 variable.

If $d.f. \geq 40$, the t-distribution can be replaced by a standard normal.

For parameters in the random part, do not use t-tests.

Simplest test for any parameters (fixed and random parts)
is the *deviance* (likelihood ratio) test, which can be used
when comparing two model fits that have used the same set of cases:
subtract deviances, use chi-squared test
($d.f.$ = number of parameters tested).

Deviance tests can be used to compare any two nested models.
If these two models do not have the same fixed parts,
then ML estimation should be used!

Other tests for parameters in the random part have been developed
which are similar to F-tests in ANOVA.

6.1 Two models with different between- and within-group regressions

	Model 1		Model 2	
Fixed Effects	Coefficient	S.E.	Coefficient	S.E.
γ_{00} = Intercept	41.15	0.23	41.15	0.23
γ_{10} = Coeff. of IQ	2.265	0.065		
γ_{20} = Coeff. of $\widetilde{\text{IQ}}$			2.265	0.065
γ_{30} = Coeff. of SES	0.161	0.011	0.161	0.011
γ_{01} = Coeff. of $\overline{\text{IQ}}$	0.647	0.264	2.912	0.262
Random Part	Parameter	S.E.	Parameter	S.E.
Level-two parameters:				
$\tau_0^2 = \text{var}(U_{0j})$	9.08	1.12	9.08	1.12
$\tau_1^2 = \text{var}(U_{1j})$	0.197	0.074	0.197	0.074
$\tau_{01} = \text{cov}(U_{0j}, U_{1j})$	–0.815	0.214	–0.815	0.214
Level-one variance:				
$\sigma^2 = \text{var}(R_{ij})$	37.42	0.91	37.42	0.91
Deviance	24661.3		24661.3	

Test for equality of within- and between-group regressions is t-test for $\overline{\text{IQ}}$ in Model 1:
$t = 0.647/0.264 = 2.45$,
$p < 0.02$.

Model 2 gives
within-group coefficient 2.265
and between-group coefficient
2.912 = 2.265 + 0.647.

However, one special circumstance: variance parameters are necessarily positive. Therefore, they may be tested one-sided.

E.g., in the random intercept model
under the null hypothesis that $\tau_0^2 = 0$,
the asymptotic distribution of –2 times the log-likelihood ratio (deviance difference) is a mixture of a point mass at 0 (with probability $\frac{1}{2}$) and a χ^2 distribution (also with probability $\frac{1}{2}$.)

The interpretation is that if the observed between-group variance is less than expected under the null hypothesis – which happens with probability $\frac{1}{2}$ – the estimate is $\hat{\tau}_0^2 = 0$ and the log-likelihood ratio is 0.

The test works as follows:
if deviance difference $= 0$, then no significance;
if deviance difference > 0, calculate p-value from χ_1^2 and divide by 2.

For testing random slope variances,
if the number of tested parameters (variances & covariances) is $p+1$,
the p-values can be obtained as
the average of the p-values for the χ^2_p and χ^2_{p+1} distributions.

(Apologies for the use of the letter p in two different meanings...)

Critical values for 50–50 mixture of χ^2_p and χ^2_{p+1} distribution.

	α			
p	0.10	0.05	0.01	0.001
1	3.81	5.14	8.27	12.81
2	5.53	7.05	10.50	15.36
3	7.09	8.76	12.48	17.61

For example: testing for a random slope in a model that further contains the random intercept but no other random slopes: $\boldsymbol{p = 1}$;
testing the second random slope: $\boldsymbol{p = 2}$;
testing the third random slope: $\boldsymbol{p = 3}$ – etc.

To test the random slope in the model of Table 5.1,
compare with Table 4.4 which is the same model but without the random slope;
deviance difference 15,227.5 – 15,213.5 $=$ 14.0.
In the table with $\boldsymbol{p = 1}$ this yields $\boldsymbol{p < 0.001}$.

Further see p. 99.

7. Explained variance

The individual variance parameters may go up when effects are added to the model.

7.1 Estimated residual variance parameters $\hat{\sigma}^2$ and $\hat{\tau}_0^2$ for models with within-group and between-group predictor variables

	$\hat{\sigma}^2$	$\hat{\tau}_0^2$
I. BALANCED DESIGN		
A. $Y_{ij} = \beta_0 + U_{0j} + E_{ij}$	8.694	2.271
B. $Y_{ij} = \beta_0 + \beta_1 \overline{X}_{.j} + U_{0j} + E_{ij}$	8.694	0.819
C. $Y_{ij} = \beta_0 + \beta_2(X_{ij} - \overline{X}_{.j}) + U_{0j} + E_{ij}$	6.973	2.443
II. UNBALANCED DESIGN		
A. $Y_{ij} = \beta_0 + U_{0j} + E_{ij}$	7.653	2.798
B. $Y_{ij} = \beta_0 + \beta_1 \overline{X}_{.j} + U_{0j} + E_{ij}$	7.685	2.038
C. $Y_{ij} = \beta_0 + \beta_2(X_{ij} - \overline{X}_{.j}) + U_{0j} + E_{ij}$	6.668	2.891

The best way to define $\boldsymbol{R^2}$, the *proportion of variance explained*, is the *proportional reduction in total variance* ;
for the random intercept model total variance is $(\boldsymbol{\sigma^2 + \tau_0^2})$.

Table 7.2 Estimating the level-1 explained variance
(balanced data)

	$\hat{\sigma}^2$	$\hat{\tau}_0^2$
A. $Y_{ij} = \beta_0 + U_{0j} + E_{ij}$	8.694	2.271
D. $Y_{ij} = \beta_0 + \beta_1(X_{ij} - \overline{X}_{.j}) + \beta_2 \overline{X}_{.j} + U_{0j} + E_{ij}$	6.973	0.991

Explained variance at level 1:

$$\boldsymbol{R_1^2 = 1 - \frac{6.973 + 0.991}{8.694 + 2.271} = 0.27.}$$

8. Heteroscedasticity

The multilevel model allows to formulate heteroscedastic models where residual variance depends on observed variables.

E.g., random part at level one = $R_{0ij} + R_{1ij}\, x_{1ij}$.

Then the level-1 variance is a quadratic function of X:

$$\text{var}(R_{0ij} + R_{1ij}\, x_{ij}) = \sigma_0^2 + 2\,\sigma_{01}\, x_{1ij} + \sigma_1^2\, x_{1ij}^2 \ .$$

For $\sigma_1^2 = 0$, this is a linear function:

$$\text{var}(R_{0ij} + R_{1ij}\, x_{ij}) = \sigma_0^2 + 2\,\sigma_{01}\, x_{1ij} \ .$$

This is possible as a variance function, without random effects interpretation.

Different statistical programs have implemented
various different variance functions.

8.1 Homoscedastic and heteroscedastic models.

	Model 1		Model 2	
Fixed Effect	Coefficient	S.E.	Coefficient	S.E.
Intercept	40.426	0.265	40.435	0.266
IQ	2.249	0.062	2.245	0.062
SES	0.171	0.011	0.171	0.011
IQ × SES	–0.020	0.005	–0.019	0.005
Gender	2.407	0.201	2.404	0.201
$\overline{\text{IQ}}$	0.769	0.293	0.749	0.292
$\overline{\text{SES}}$	–0.093	0.042	–0.091	0.042
$\overline{\text{IQ}} \times \overline{\text{SES}}$	–0.105	0.033	–0.107	0.033
Random Part	Parameters	S.E.	Parameters	S.E.
Level-two random part:				
Intercept variance	8.321	1.036	8.264	1.030
IQ slope variance	0.146	0.065	0.146	0.065
Intercept - IQ slope covariance	–0.898	0.197	–0.906	0.197
Level-one variance:				
σ_0^2 constant term	35.995	0.874	37.851	1.280
σ_{01} gender effect			–1.887	0.871
Deviance	24486.8		24482.2	

This shows that there is significant evidence for heteroscedasticity:
$\chi^2_1 = 4.6,\ p < 0.05$.

The estimated residual (level-1) variance is
37.85 for boys and 37.85 – 2×1.89 = 34.07 for girls.

The following models show, however, that the heteroscedasticity as a function of IQ is more important.
First look only at Model 3.

8.2 Heteroscedastic models depending on IQ.

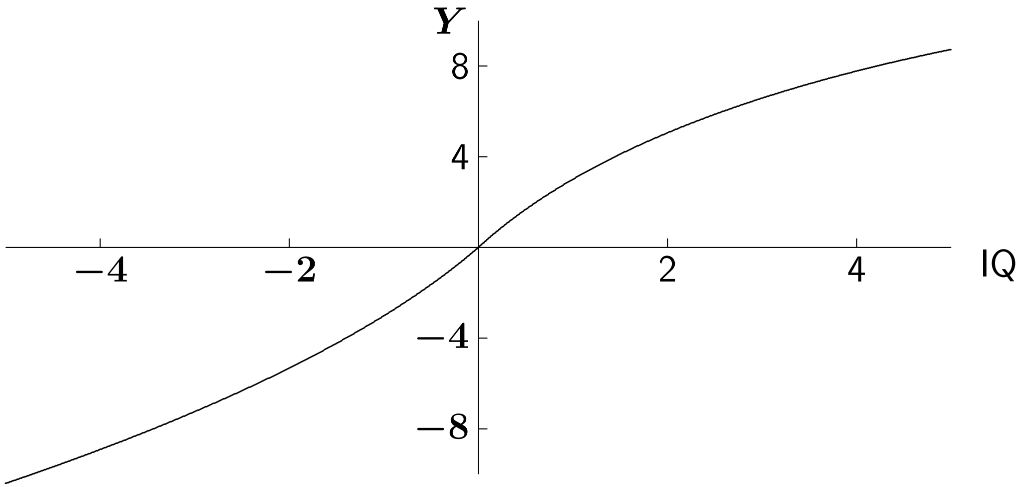
	Model 3		Model 4	
Fixed Effect	Coefficient	S.E.	Coefficient	S.E.
Intercept	40.51	0.26	40.51	0.27
IQ	2.200	0.058	3.046	0.125
SES	0.175	0.011	0.168	0.011
IQ × SES	–0.022	0.005	–0.016	0.005
Gender	2.311	0.198	2.252	0.196
$\overline{\text{IQ}}$	0.685	0.289	0.800	0.284
$\overline{\text{SES}}$	–0.087	0.041	–0.083	0.041
$\overline{\text{IQ}} \times \overline{\text{SES}}$	–0.107	0.033	–0.089	0.032
IQ_{-}^{2}			0.193	0.038
IQ_{+}^{2}			–0.260	0.033
Random Part	Parameter	S.E.	Parameter	S.E.
Level-two random effects:				
Intercept variance	8.208	1.029	7.989	1.002
IQ slope variance	0.108	0.057	0.044	0.048
Intercept - IQ slope covariance	–0.733	0.187	–0.678	0.171
Level-one variance parameters:				
σ_0^2 constant term	36.382	0.894	36.139	0.887
σ_{01} IQ effect	–1.689	0.200	–1.769	0.191
Deviance	24430.2		24369.0	

The level-1 variance function for Model 3 is 36.38 – 3.38 IQ .

Maybe further differentiation is possible between low-IQ pupils?
Model 4 uses

$$\text{IQ}^2_- = \begin{cases} \text{IQ}^2 & \text{if IQ} < 0 \\ 0 & \text{if IQ} \geq 0 \,, \end{cases}$$

$$\text{IQ}^2_+ = \begin{cases} 0 & \text{if IQ} < 0 \\ \text{IQ}^2 & \text{if IQ} \geq 0 \,. \end{cases}$$



Effect of IQ on language test as estimated by Model 4.

Heteroscedasticity can be very important for the researcher (although mostly she/he doesn't know it yet).

Bryk & Raudenbush: Correlates of diversity.
Explain not only means, but also variances!

Heteroscedasticity also possible for level-2 random effects:
give a random slope at level 2 to a level-2 variable.

10. Assumptions of the Hierarchical Linear Model

$$\boldsymbol{Y_{ij}} = \gamma_0 + \sum_{h=1}^{r} \gamma_h \, \boldsymbol{x_{hij}} + \boldsymbol{U_{0j}} + \sum_{h=1}^{p} \boldsymbol{U_{hj}} \, \boldsymbol{x_{hij}} + \boldsymbol{R_{ij}} \; .$$

Questions:

1. Does the fixed part contain the right variables (now $\boldsymbol{X_1}$ to $\boldsymbol{X_r}$)?
2. Does the random part contain the right variables (now $\boldsymbol{X_1}$ to $\boldsymbol{X_p}$)?
3. Are the level-one residuals normally distributed?
4. Do the level-one residuals have constant variance?
5. Are the level-two random coefficients normally distributed with mean 0?
6. Do the level-two random coefficients have a constant covariance matrix?

Follow the logic of the HLM

1. Include contextual effects

For every level-1 variable $\boldsymbol{X_h}$, check the fixed effect of the group mean $\boldsymbol{\bar{X}_h}$.

Econometricians' wisdom: "the $\boldsymbol{U_{0j}}$ must not be correlated with the $\boldsymbol{X_{hij}}$.
Therefore test this correlation by testing the effect of $\boldsymbol{\bar{X}_h}$ ('Hausman test')
Use a fixed effects model if this effect is significant".

Different approach to the same assumption:
Include the fixed effect of $\boldsymbol{\bar{X}_h}$ if it is significant,
and continue to use a random effects model.
(Also check effects of variables $\boldsymbol{\bar{X}_{h.j}\, Z_j}$ for cross-level interactions involving $\boldsymbol{X_h}$!)

Also the random slopes $\boldsymbol{U_{hj}}$ must not be correlated with the $\boldsymbol{X_{kij}}$.
This can be checked by testing the fixed effect of $\boldsymbol{\bar{X}_{k.j}\, X_{hij}}$.

This procedure widens the scope of random coefficient models beyond what is allowed by the conventional rules of econometricians.

Assumption that level-2 random effects $\boldsymbol{U}_j$ have zero means.

What kind of bias can occur if this assumption is made but does not hold?

For a misspecified model,
suppose that we are considering a random intercept model:

$$\boldsymbol{Z}_j = \boldsymbol{1}_j$$

where the expected value of $\boldsymbol{U}_j$ is not 0 but

$$E\boldsymbol{U}_j = \boldsymbol{z}_{2j}\,\gamma_\star$$

for $\boldsymbol{1} \times \boldsymbol{r}$ vectors $\boldsymbol{z}_{2j}$ and an unknown regression coefficient $\gamma_\star$. Then

$$\boldsymbol{U}_j = \boldsymbol{z}_{2j}\,\gamma_\star + \tilde{\boldsymbol{U}}_j$$

with

$$E\tilde{\boldsymbol{U}}_j = \boldsymbol{0}\ .$$

Write $\boldsymbol{X}_j = \bar{\boldsymbol{X}}_j + \tilde{\boldsymbol{X}}_j$, where $\bar{\boldsymbol{X}}_j = \mathbf{1}_j\,(\mathbf{1}_j'\mathbf{1}_j)^{-1}\mathbf{1}_j'\,\boldsymbol{X}_j$ are the group means.
Then the data generating mechanism is

$$\boldsymbol{Y}_j = \bar{\boldsymbol{X}}_j\,\gamma + \tilde{\boldsymbol{X}}_j\,\gamma + \mathbf{1}_j\,\boldsymbol{z}_{2j}\,\gamma_\star + \mathbf{1}_j\,\tilde{\boldsymbol{U}}_j + \boldsymbol{R}_j\ ,$$

where $E\tilde{\boldsymbol{U}}_j = \mathbf{0}$.
There will be a bias in the estimation of γ
if the matrices $\boldsymbol{X}_j = \bar{\boldsymbol{X}}_j + \tilde{\boldsymbol{X}}_j$ and $\mathbf{1}_j\,\tilde{\boldsymbol{U}}_j$ are not orthogonal.
By construction, $\tilde{\boldsymbol{X}}_j$ and $\mathbf{1}_j\,\tilde{\boldsymbol{U}}_j$ are orthogonal, so the difficulty is with $\bar{\boldsymbol{X}}_j$.

The solution is to give $\bar{\boldsymbol{X}}_j$ and $\tilde{\boldsymbol{X}}_j$ separate effects:

$$\boldsymbol{Y}_j = \bar{\boldsymbol{X}}_j\,\gamma_1 + \tilde{\boldsymbol{X}}_j\,\gamma_2 + \mathbf{1}_j\boldsymbol{U}_j + \boldsymbol{R}_j\ .$$

Now γ_2 has the role of the old γ:
'the estimation is done using only within-group information'.

Often, there are substantive interpretations of the difference between the *within-group effects* γ_2 and the *between-group effects* γ_1.

2. Check random effects of level-1 variables.

See Chapter 5.

4. Check heteroscedasticity.

See Chapter 8.

3,4. Level-1 residual analysis

5,6. Level-2 residual analysis

For residuals in multilevel models, more information is in Chapter 3 of *Handbook of Multilevel Analysis* (eds. De Leeuw and Meijer, Springer 2008) (preprint at course website).

Level-one residuals

OLS within-group residuals can be written as

$$\hat{\boldsymbol{R}}_j = \left(\boldsymbol{I}_{n_j} - \boldsymbol{P}_j\right) \boldsymbol{Y}_j$$

where we define design matrices $\check{\boldsymbol{X}}_j$ comprising $\boldsymbol{X}_j$ as well as $\boldsymbol{Z}_j$ (to the extent that $\boldsymbol{Z}_j$ is not already included in $\boldsymbol{X}_j$) and

$$\boldsymbol{P}_j = \check{\boldsymbol{X}}_j(\check{\boldsymbol{X}}_j'\check{\boldsymbol{X}}_j)^{-1}\check{\boldsymbol{X}}_j' \, .$$

Model definition implies

$$\hat{\boldsymbol{R}}_j = \left(\boldsymbol{I}_{n_j} - \boldsymbol{P}_j\right) \boldsymbol{R}_j \quad :$$

these level-1 residuals are not confounded by $\boldsymbol{U}_j$.

Use of level-1 residuals :

Test the fixed part of the level-1 model using OLS level-1 residuals, calculated per group separately.

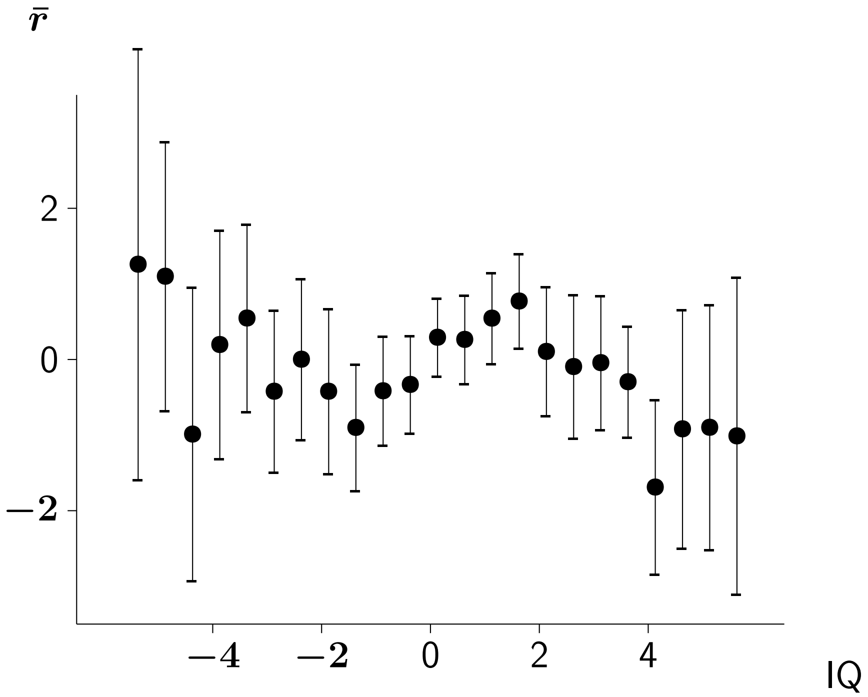
Test the random part of the level-1 model using squared standardized OLS residuals.

In other words, the level-1 specification can be studied by disaggregation to the within-group level (comparable to a "fixed effects analysis").

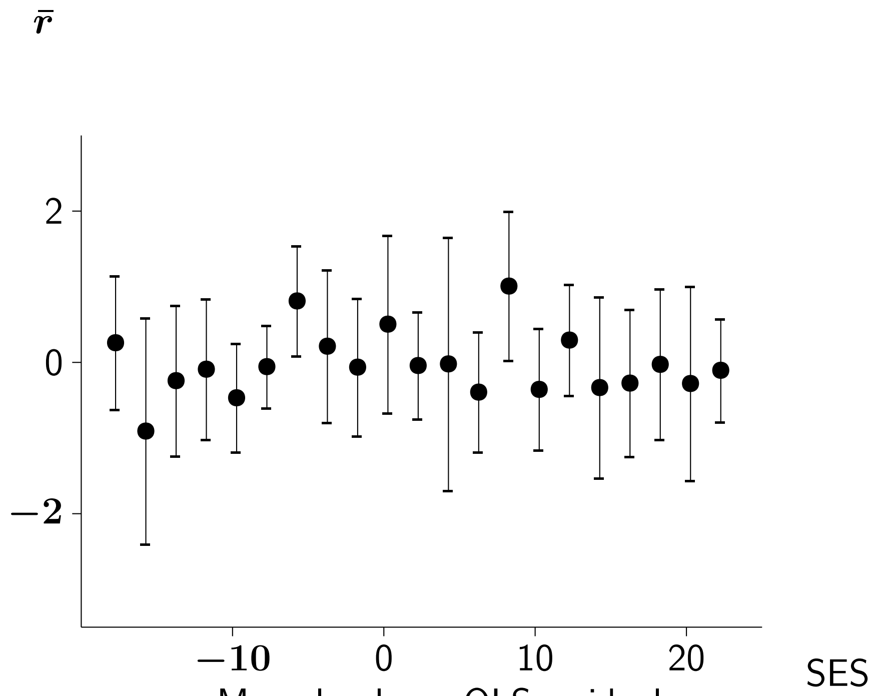
The construction of the OLS within-group residuals implies that this tests only the level-one specification, independent of the correctness of the level-two specification.

The examples of Chapter 8 are taken up again.

Example: model with effects of IQ, SES, sex.



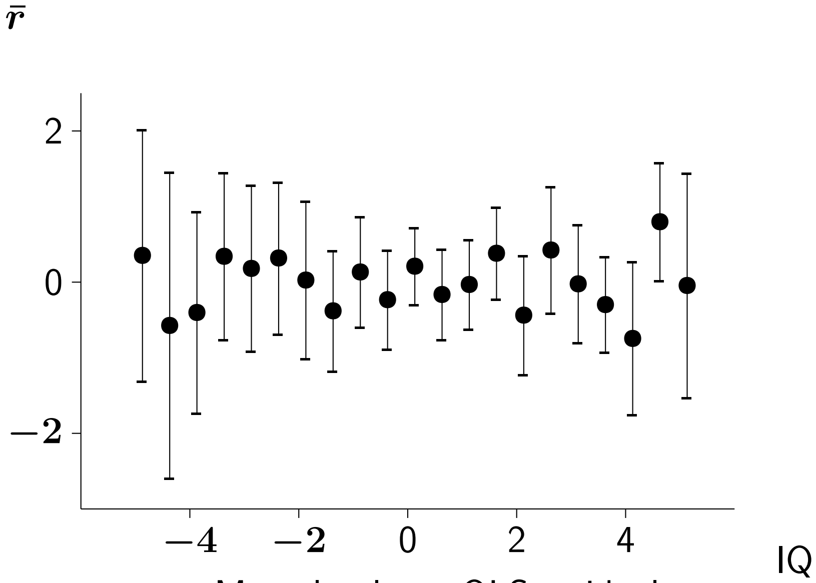
Mean level-one OLS residuals
(bars ∼ twice standard error of the mean)
as function of IQ.



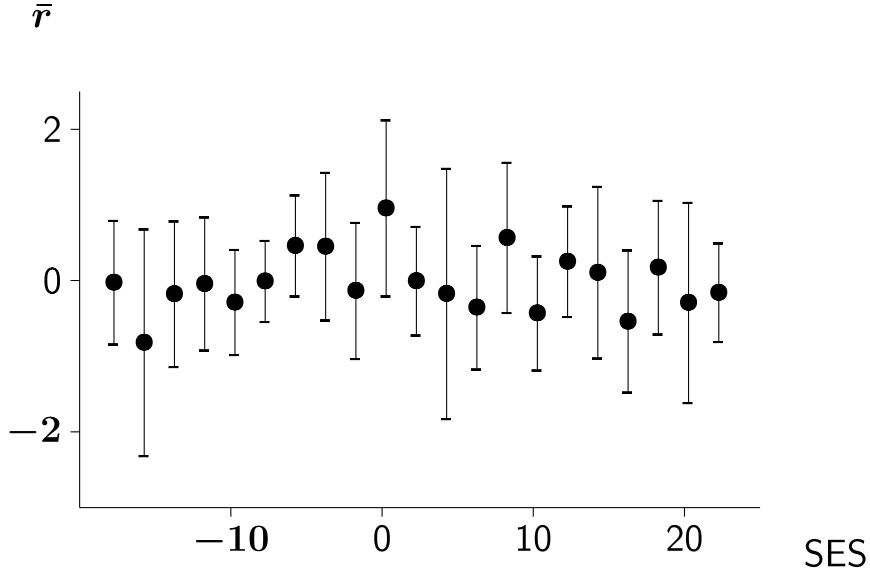
Mean level-one OLS residuals
as function of SES.

This suggest a curvilinear effect of IQ.

Model with effects also of IQ_{-}^{2} and IQ_{+}^{2} .



Mean level-one OLS residuals as function of IQ.



Mean level-one OLS residuals as function of SES.

This looks pretty random.

Are the within-group residuals normally distributed?

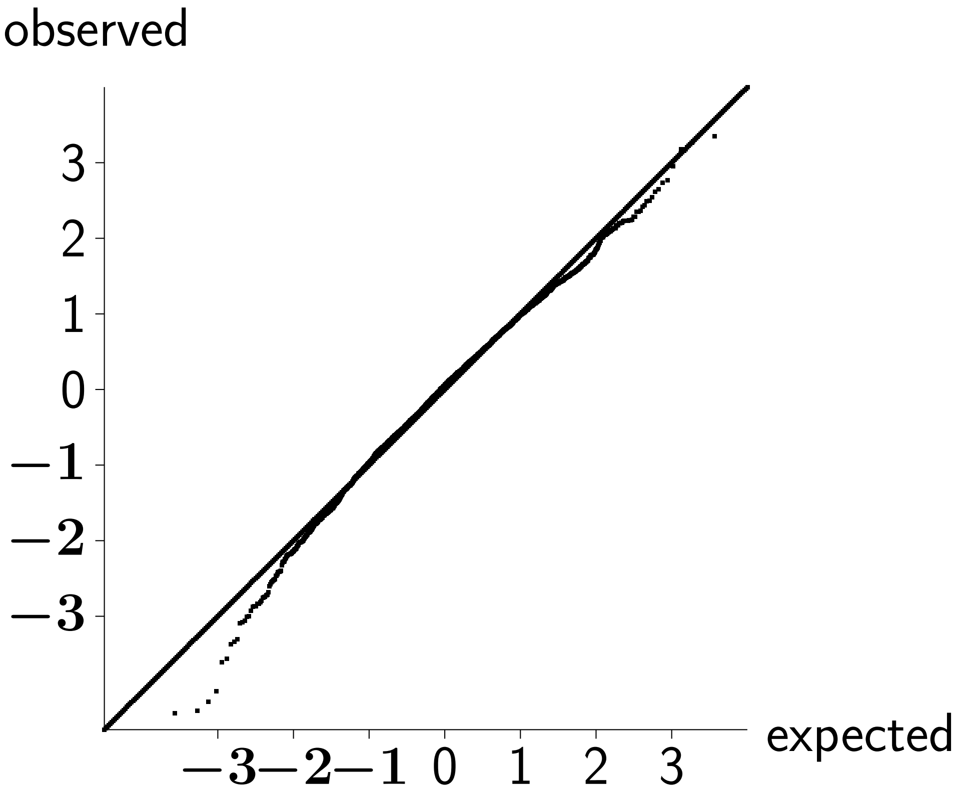


Figure 10.3 Normal probability plot of standardized level-one OLS residuals.

Left tail is a bit heavy, but this is not serious.

Level-two residuals

Empirical Bayes (EB) level-two residuals defined as conditional means

$$\hat{\boldsymbol{U}}_j = E\{\boldsymbol{U}_j \mid \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$$

(using parameter estimates $\hat{\gamma}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\xi}}$)

$$= \hat{\boldsymbol{\Omega}}\, \boldsymbol{Z}_j'\, \hat{\boldsymbol{V}}_j^{-1}\, (\boldsymbol{Y}_j - \boldsymbol{X}_j \hat{\gamma}_j) = \hat{\boldsymbol{\Omega}}\, \boldsymbol{Z}_j'\, \hat{\boldsymbol{V}}_j^{-1}\, (\boldsymbol{Z}_j\, \boldsymbol{U}_j \, + \, \boldsymbol{R}_j - \boldsymbol{X}_j(\hat{\gamma} - \gamma))$$

where

$$\boldsymbol{V}_j = \mathit{Cov}\, \boldsymbol{Y}_j = \boldsymbol{Z}_j \boldsymbol{\Omega} \boldsymbol{Z}_j' \, + \, \boldsymbol{\Sigma}_j \; , \; \hat{\boldsymbol{V}}_j = \boldsymbol{Z}_j \hat{\boldsymbol{\Omega}} \boldsymbol{Z}_j' \, + \, \hat{\boldsymbol{\Sigma}}_j \; ,$$

with $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\hat{\boldsymbol{\xi}})$ and $\hat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}_j(\hat{\boldsymbol{\theta}})$.

You don't need to worry about the formulae.

'Diagnostic variances', used for assessing distributional properties of $\boldsymbol{U_j}$:

$$\mathit{Cov}\, \hat{\boldsymbol{U}}_j \approx \boldsymbol{\Omega Z_j' V_j^{-1} Z_j \Omega} \ ,$$

'Comparative variances', used for comparing 'true values' $\boldsymbol{U_j}$ of groups:

$$\mathit{Cov}\ \left(\hat{\boldsymbol{U}}_j - \boldsymbol{U}_j\right) \approx \boldsymbol{\Omega} - \boldsymbol{\Omega Z_j' V_j^{-1} Z_j \Omega} \ .$$

Note that

$$\mathit{Cov}\ (\boldsymbol{U}_j) = \mathit{Cov}\ (\boldsymbol{U}_j - \hat{\boldsymbol{U}}_j) \ + \ \mathit{Cov}\ (\hat{\boldsymbol{U}}_j) \ .$$

Standardization (by diagnostic variances) :

$\sqrt{\hat{\boldsymbol{U}}_j' \{\widehat{\mathit{Cov}}\ (\hat{\boldsymbol{U}}_j)\}^{-1} \hat{\boldsymbol{U}}_j}$ (with the sign reinstated)
is the standardized EB residual.

However,

$$\hat{\boldsymbol{U}}_j'\{\widehat{\mathit{Cov}}\,(\hat{\boldsymbol{U}}_j)\}^{-1}\hat{\boldsymbol{U}}_j \approx \hat{\boldsymbol{U}}_j^{(\mathsf{OLS})\prime}\left(\hat{\boldsymbol{\sigma}}^2(\boldsymbol{Z}_j'\boldsymbol{Z}_j)^{-1} + \hat{\boldsymbol{\Omega}}\right)^{-1}\hat{\boldsymbol{U}}_j^{(\mathsf{OLS})}$$

where $\hat{\boldsymbol{U}}_j^{(\mathsf{OLS})} = (\boldsymbol{Z}_j'\boldsymbol{Z}_j)^{-1}\boldsymbol{Z}_j'\,(\boldsymbol{Y}_j - \boldsymbol{X}_j\hat{\gamma}_j)$

is the OLS estimate of $\boldsymbol{U}_j$, estimated from level-1 residuals $\boldsymbol{Y}_j - \boldsymbol{X}_j\hat{\gamma}_j$.

This shows that standardization by diagnostic variances
takes away the difference between OLS and EB residuals.
Therefore, in checking standardized level-two residuals,
the distinctoin between OLS and EB residuals loses its meaning.

Test the fixed part of the level-2 model using non-standardized EB residuals.

Test the random part of the level-2 model
using squared EB residuals standardized by *diagnostic* variance.

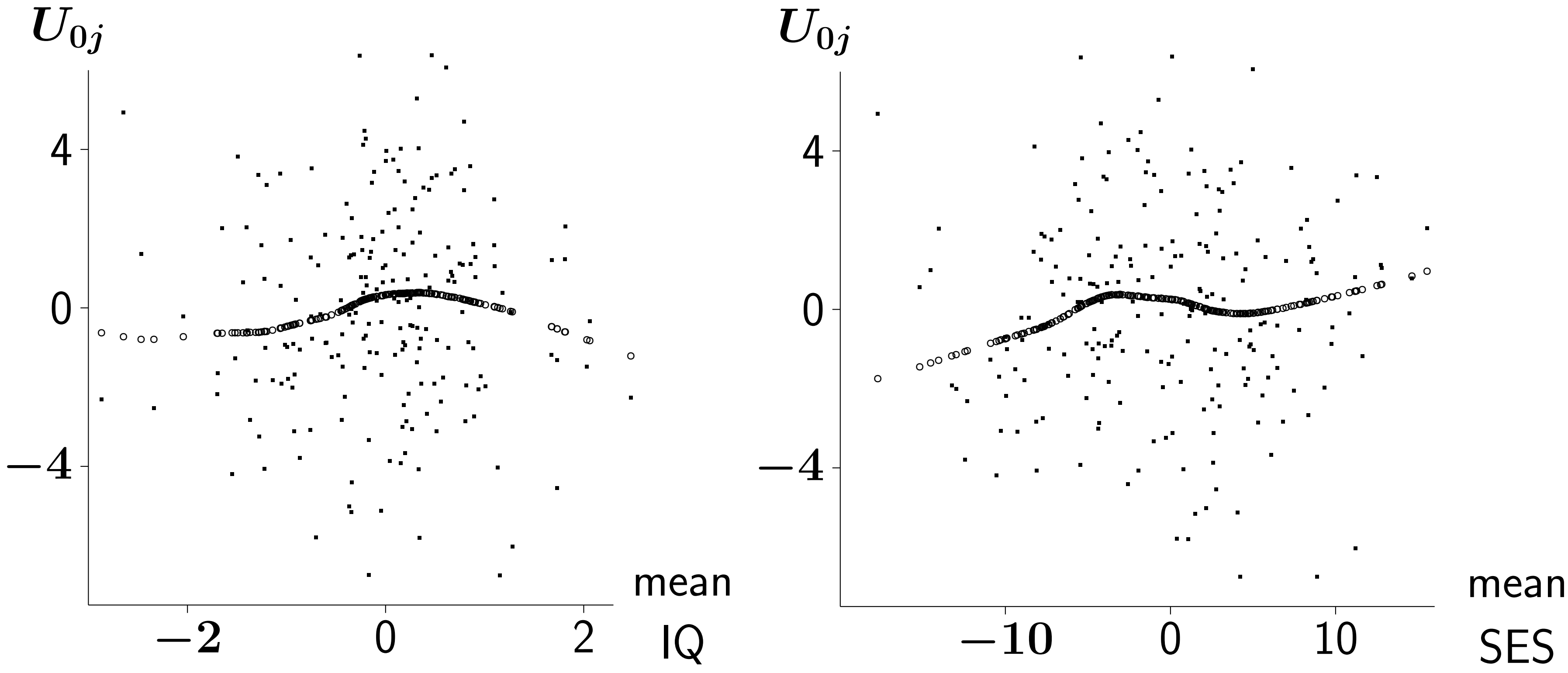


Figure 10.4 Posterior intercepts as function of (left) average IQ and (right) average SES per school. Smooth lowess approximations are indicated by ∘.

The slight deviations do not lead to concerns.

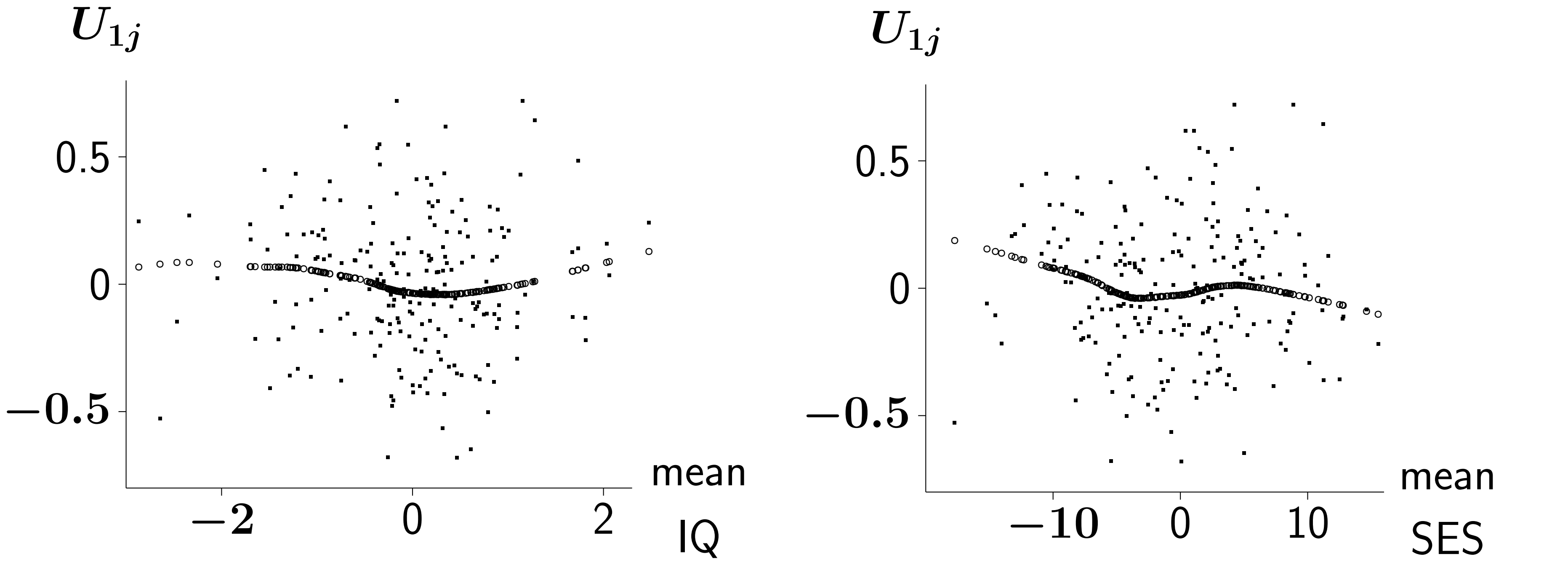


Figure 10.5 Posterior IQ slopes as function of (left) average IQ and (right) average SES per school. Smooth lowess approximations are indicated by ◦.

Again, the slight deviations do not lead to concerns.

Multivariate residuals

The multivariate residual is defined, for level-two unit $\boldsymbol{j}$, as

$$\boldsymbol{Y}_j - \boldsymbol{X}_j \hat{\gamma}.$$

The standardized multivariate residual is defined as

$$\boldsymbol{M}_j^2 = (\boldsymbol{Y}_j - \boldsymbol{X}_j\,\hat{\gamma}_j)'\,\hat{\boldsymbol{V}}_j^{-1}\,(\boldsymbol{Y}_j - \boldsymbol{X}_j\,\hat{\gamma}_j)\,.$$

If all variables with fixed effects also have random effects, then

$$\boldsymbol{M}_j^2 = (\boldsymbol{n}_j - \boldsymbol{t}_j)\,\boldsymbol{s}_j^2 \,+\, \hat{\boldsymbol{U}}_j'\,\{\widehat{\mathit{Cov}}\,(\hat{\boldsymbol{U}}_j)\}^{-1}\,\hat{\boldsymbol{U}}_j\ ,$$

where

$$\boldsymbol{s}_j^2 = \frac{1}{\boldsymbol{n}_j - \boldsymbol{t}_j}\,\hat{\boldsymbol{R}}_j'\,\hat{\boldsymbol{R}}_j\ ,\ \boldsymbol{t}_j = \mathrm{rank}(\boldsymbol{X}_j)\ .$$

This indicates how well the model fits to group $\boldsymbol{j}$.
Note the confounding with level-1 residuals.

If an ill-fitting group does not have a strong effect on the parameter estimates, then it is not so serious.

Deletion residuals

The deletion standardized multivariate residual can be used to assess the fit of group j, but takes out the effect of this group on the parameter estimates:

$$M^2_{(-j)} = \left(Y_j - X_j\,\hat{\gamma}_{(-j)}\right)'\,\hat{V}^{-1}_{(-j)}\left(Y_j - X_j\,\hat{\gamma}_{(-j)}\right)$$

where

$$\hat{V}_{(-j)} = Z_j\,\hat{\Omega}_{(-j)}\,Z'_j \,+\, \hat{\Sigma}_{(-j)}\ ,$$

$_{(-j)}$ meaning that group j is deleted from the data for estimating this parameter.

Full computation of deletion estimates may be computing-intensive,
which is unattractive for diagnostic checks.
Approximations have been proposed:
Lesaffre & Verbeke: Taylor series; Snijders & Bosker: one-step estimates.
However, with fast computers
full computation of deletion estimates now also is feasible.

The approximate distribution of multivariate residuals, if the model fits well and sample sizes are large, is χ^2, d.f. $= n_j$.

Influence diagnostics of higher-level units

The *influence* of the groups can be assessed by statistics analogous to Cook's distance:
how large is the influence of this group on the parameter estimates?

Standardized measures of influence of unit j on fixed parameter estimates :

$$C_j^F = \frac{1}{r} \left(\hat{\gamma} - \hat{\gamma}_{(\text{-}j)}\right)' \hat{S}_{F(\text{-}j)}^{-1} \left(\hat{\gamma} - \hat{\gamma}_{(\text{-}j)}\right)$$

where S_F is covariance matrix of fixed parameter estimates, and $_{(\text{-}j)}$ means that group j is deleted from the data for estimating this parameter.

on random part parameters :

$$C_j^R = \frac{1}{p} \left(\hat{\eta} - \hat{\eta}_{(\text{-}j)}\right)' \hat{S}_{R(\text{-}j)}^{-1} \left(\hat{\eta} - \hat{\eta}_{(\text{-}j)}\right) \; ,$$

combined :

$$C_j = \frac{1}{r + p} \left(r C_j^F + p C_j^R\right) .$$

Values of C_j larger than 1 indicate strong outliers.
Values larger than $4/N$ may merit inspection.

Table 10.1 the 20 largest influence statistics, and p-values for multivariate residuals,
of the 211 schools; Model 4 of Chapter 8 but without heteroscedasticity.

School	n_j	C_j	p_j
182	9	0.053	0.293
107	17	0.032	0.014
229	9	0.028	0.115
14	21	0.027	0.272
218	24	0.026	0.774
52	21	0.025	0.024
213	19	0.025	0.194
170	27	0.021	0.194
67	26	0.017	0.139
18	24	0.016	0.003

School	n_j	C_j	p_j
117	27	0.014	0.987
153	22	0.013	0.845
187	26	0.013	0.022
230	21	0.012	0.363
15	8	0.012	0.00018
256	10	0.012	0.299
122	23	0.012	0.005
50	24	0.011	0.313
101	23	0.011	0.082
214	21	0.011	0.546

School 15 does not survive Bonferroni correction: $211 \times 0.00018 = 0.038$.
Therefore now add the heteroscedasticity of Model 4 in Chapter 8.

Table 10.2 the 20 largest influence statistics, and p-values for multivariate residuals,
of the 211 schools; Model 4 of Chapter 8 with heteroscedasticity.

School	n_j	C_j	p_j
213	19	0.094	0.010
182	9	0.049	0.352
107	17	0.041	0.006
187	26	0.035	0.009
52	21	0.028	0.028
218	24	0.025	0.523
14	21	0.024	0.147
229	9	0.016	0.175
67	26	0.016	0.141
122	23	0.016	0.004

School	n_j	C_j	p_j
18	24	0.015	0.003
230	21	0.015	0.391
169	30	0.014	0.390
170	27	0.013	0.289
144	16	0.013	0.046
117	27	0.013	0.988
40	25	0.012	0.040
153	22	0.012	0.788
15	8	0.011	0.00049
202	14	0.010	0.511

School 15 now does survive the Bonferroni correction: $211 \times 0.00049 = 0.103$.
Therefore now add the heteroscedasticity of Model 4 in Chapter 8.
Another school (108) does have poor fit $\boldsymbol{p = 0.00008}$, but small influence ($\boldsymbol{C_j = 0.008}$).

Leaving out ill-fitting schools does not lead to appreciable differences in results.

The book gives further details.

11. Designing Multilevel Studies

Note: each level corresponds to a sample from a population.
For each level, the total sample size counts.

E.g., in a 3-level design:
15 municipalities;
in each municipality, 200 households;
in each household, 2-4 individuals.

Total sample sizes are 15 (level 3), 3000 (level 2), $\approx$ 9000 (level 1).
Much information about individuals and households, but little about municipalities.

Power and standard errors

Consider testing a parameter β , based on a t-ratio

$$\frac{\hat{\beta}}{\text{s.e.}(\hat{\beta})} \, .$$

For significance level α and power γ

$$\frac{\beta}{s.e.(\hat{\beta})} \approx (z_{1-\alpha} + z_{\gamma}) = (z_{1-\alpha} - z_{1-\gamma}) \; ,$$

where $z_{1-\alpha}$, z_{γ} and $z_{1-\gamma}$ are values for standard normal distribution.
E.g., for $\alpha = .05,\ \gamma = .80$, effect size $\beta = .20$, sample size must be such that

$$\text{standard error} \leq \frac{.20}{1.64 + 0.84} = 0.081 \, .$$

The following discussion is mainly in terms of standard errors,
always for two-level designs,
with two-stage samples of N clusters each with n units.

Design effect for estimation of a mean

Empty model:

$$Y_{ij} = \mu + U_j + R_{ij} .$$

with $\text{var}(U_j) = \tau^2$, $\text{var}(R_{ij}) = \sigma^2$.
Parameter to be estimated is μ.

$$\hat{\mu} = \frac{1}{Nn} \sum_{j=1}^{N} \sum_{i=1}^{n} Y_{ij} ,$$

$$\text{var}(\hat{\mu}) = \frac{\tau^2}{N} + \frac{\sigma^2}{Nn} .$$

The sample mean of a simple random sample of Nn elements from this population has variance

$$\frac{\tau^2 + \sigma^2}{Nn} .$$

The relative efficiency of simple random sample w.r.t. two-stage sample is the *design effect* (cf. Kish, Cochran)

$$\frac{n\tau^2 + \sigma^2}{\tau^2 + \sigma^2} = 1 + (n-1)\rho_{\mathrm{I}} \ , \qquad (1)$$

where

$$\rho_{\mathrm{I}} = \frac{\tau^2}{\tau^2 + \sigma^2} \, .$$

Effect of a level-two variable

Two-level regression with random intercept model

$$Y_{ij} = \beta_0 + \beta_1 x_j + U_j + E_{ij} .$$

When $\text{var}(X) = s_X^2$,

$$\text{var}(\hat{\beta}_1) = \frac{\tau^2 + (\sigma^2/n)}{N s_X^2} .$$

For a simple random sample from the same population,

$$\text{var}(\hat{\beta}_1^{\text{disaggregated}}) = \frac{\tau^2 + \sigma^2}{N n s_X^2} .$$

Relative efficiency again equal to (1).

Effect of a level-one variable

Now suppose X is a pure level-one variable,
i.e., $\bar{X}$ the same in each cluster, $\rho_{\mathrm{I}} = -1/(n-1)$.
Assume var$(X) = s_X^2$ within each cluster.
E.g.: time effect in longitudinal design; within-cluster randomization.
Again

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + U_j + E_{ij} .$$

Now

$$\hat{\beta}_1 = \frac{1}{N n s_X^2} \sum_{j=1}^{N} \sum_{i=1}^{n} x_{ij} Y_{ij} \qquad (2)$$

$$= \beta_1 + \frac{1}{N n s_X^2} \sum_{j=1}^{N} \sum_{i=1}^{n} x_{ij} E_{ij}$$

with variance

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{N n s_X^2} .$$

For a simple random sample from the same population,

$$\text{var}(\hat{\beta}_1^{\text{disaggregated}}) = \frac{\sigma^2 + \tau^2}{N n s_X^2} .$$

The design effect now is

$$\frac{\sigma^2}{\tau^2 + \sigma^2} = 1 - \rho_\mathsf{I} \; .$$

Note the efficiency due to blocking on clusters!

For random intercept models,
design effect < 1 for level-one variables,
and > 1 for level-two variables.

Conclusion: for a comparison of randomly assigned treatments
with costs depending on Nn,
randomising within clusters is more efficient than between clusters
(Moerbeek, van Breukelen, and Berger, JEBS 2000.)

But not necessarily for variables with a random slope!

Level-one variable with random slope

Assume a random slope for X:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + U_{0j} + U_{1j}\, x_{ij} + E_{ij}\,.$$

The variance of (2) now is

$$\text{var}(\hat{\beta}_1) = \frac{n\tau_1^2 s_X^2 + \sigma^2}{Nns_X^2}\,.$$

The marginal residual variance of Y is

$$\sigma^2 + \tau_0^2 + \tau_1^2 s_X^2$$

so the design effect now is

$$\frac{n\tau_1^2 s_X^2 + \sigma^2}{\tau_0^2 + \tau_1^2 s_X^2 + \sigma^2}\,.$$

A two-stage sample (like the "within-subject design" in psychology)
'neutralizes' variability due to random intercept, not due to random slope of $\boldsymbol{X}$.

In practice:
if there is a random slope for $\boldsymbol{X}$ then the fixed effect does not tell the whole story and it is relevant to choose a design in which the slope variance can be estimated.

Optimal sample size for estimating a regression coefficient

Estimation variance of regression parameters ≈

$$\frac{\sigma_1^2}{N} + \frac{\sigma_2^2}{Nn}$$

for suitable σ_1^2, σ_2^2 .

Total cost is usually not a function of total sample size Nn, but of the form

$$c_1 N + c_2 N n \ .$$

Minimizing the variance under the constraint of a given total budget
leads to the optimum

$$n_{\text{opt}} = \sqrt{\frac{c_1 \sigma_2^2}{c_2 \sigma_1^2}}$$

(which still must be rounded).
Optimal N depends on available budget.

For level-one variables with constant cluster means,

$$\sigma_1^2 = 0$$

so that $n_{\text{opt}} = \infty$: the single-level design.

For level-two variables,

$$\sigma_1^2 = \frac{\tau^2}{s_X^2} \quad \sigma_2^2 = \frac{\sigma^2}{s_X^2}$$

so that

$$n_{\text{opt}} = \sqrt{\frac{c_1 \sigma^2}{c_2 \tau^2}} .$$

Cf. Cochran (1977),

Snijders and Bosker (JEBS 1993),

Raudenbush (Psych. Meth. 1997, p. 177),

Chapter Snijders in Leyland & Goldstein (eds., 2001),

Moerbeek, van Breukelen, and Berger (JEBS 2000),

Moerbeek, van Breukelen, and Berger (The Statistician, 2001).

How much power is gained by using covariates?

In single-level regression:
covariate reduces unexplained variance by factor $1 - \rho^2$.

Assume random intercept models:

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 x_{ij} + U_j + R_{ij} \\ Y_{ij} &= \tilde{\beta}_0 + \beta_1 x_{ij} + \beta_2 z_{ij} + \tilde{U}_j + \tilde{R}_{ij} \\ Z_{ij} &= \gamma_0 + U_{Zj} + R_{Zij} \, . \end{aligned}$$

Also assume that Z is uncorrelated with X within and between groups (regression coefficient β_1 not affected by control for Z).

Denote the population residual within-group correlation between $\boldsymbol{Y}$ and $\boldsymbol{Z}$ by

$$\boldsymbol{\rho_W = \rho(R_{ij},\, R_{Zij})},$$

and the population residual between-group correlation by

$$\boldsymbol{\rho_B = \rho(U_j,\, U_{Zj})}\ .$$

The reduction in variance parameters is given by

$$\begin{aligned}\tilde{\boldsymbol{\sigma}}^2 &= (1 - \boldsymbol{\rho}_W^2)\,\boldsymbol{\sigma}^2\,,\\ \tilde{\boldsymbol{\tau}}^2 &= (1 - \boldsymbol{\rho}_B^2)\,\boldsymbol{\tau}^2\ .\end{aligned}$$

In formulae for standard errors, control for $\boldsymbol{Z}$
leads to replacement of $\boldsymbol{\sigma}^2$ and $\boldsymbol{\tau}^2$ by $\tilde{\boldsymbol{\sigma}}^2$ and $\tilde{\boldsymbol{\tau}}^2$.

Therefore:
for pure within-group variables, only within-group correlation counts;
for level-two variables, both correlations count,
but between-group correlations play the major role unless $\boldsymbol{n}$ is rather small.

Estimating fixed effects in general

The assumptions in the preceding treatment are very stringent.

Approximate expressions for standard errors were given by Snijders and Bosker (JEBS 1993)
and implemented in computer program *PinT* ('Power in Two-level designs')
available from

http://www.stats.ox.ac.uk/~snijders/multilevel.htm

Raudenbush (Psych. Meth. 1997) gives more detailed formulae
for some special cases.

Main difficulty in the practical application is the necessity
to specify parameter values:
means,
covariance matrices of explanatory variables,
random part parameters.

Example:
sample sizes for therapy effectiveness study.

Outcome variable $\boldsymbol{Y}$, unit variance, depends on
$\boldsymbol{X_1}$ (0–1) course for therapists: study variable,
$\boldsymbol{X_2}$ therapists' professional training,
$\boldsymbol{X_3}$ pretest.

Suppose pre-thinking leads to the following guesstimates:

Means: $\boldsymbol{\mu_1 = 0.4,\ \mu_2 = \mu_3 = 0}$.
Variances between groups:
var$(\boldsymbol{X_1}) = \mathbf{0.24}$ (because $\boldsymbol{\mu_1 = 0.4}$)
var$(\boldsymbol{X_2})$ = var$(\boldsymbol{X_3}) = \mathbf{1}$ (standardization)
$\boldsymbol{\rho}_{\mathrm{I}}\ (\boldsymbol{X_3}) = \mathbf{0.19}$ (prior knowledge)
$\boldsymbol{\rho(X_1, X_2) = -0.4}$ (conditional randomization)
$\boldsymbol{\rho(X_1, \bar{X}_3 \mid X_2) = 0}$ (randomization) $\Rightarrow \boldsymbol{\rho(X_1, \bar{X}_3) = 0.2}$
$\boldsymbol{\rho(X_2, \bar{X}_3) = 0.5}$ (prior knowledge).

This yields $\boldsymbol{\sigma^2_{X_3(W)} = 1 - 0.19 = 0.81}$ and

$$\boldsymbol{\Sigma_{X(B)} = \begin{pmatrix} 0.24 & -0.20 & 0.04 \\ -0.20 & 1.0 & 0.22 \\ 0.04 & 0.22 & 0.19 \end{pmatrix}} \; .$$

Parameters of the random part:

$$\text{var}(Y_{ij}) = \beta_1^2 \sigma^2_{X(W)} + \beta' \Sigma_{X(B)} \beta + \tau_0^2 + \sigma^2$$

(cf. Section 7.2 of Snijders and Bosker, 2012).
Therefore total level-one variance of Y is

$$\beta_1^2 \sigma_{X(W)} + \sigma^2$$

and total level-two variance is

$$\beta' \Sigma_{X(B)} \beta + \tau_0^2 \; .$$

Suppose that $\rho_{\text{I}}\,(Y) = 0.2$ and that the available explanatory variables together explain 0.25 of the level-one variance and 0.5 of the level-two variance.
Then $\sigma^2 = 0.6$ and $\tau_0^2 = 0.10$.
Budget structure:
assume $c_1 = 20$, $c_2 = 1$, $k = 1000$.

With this information, $PinT$ can run. Results:

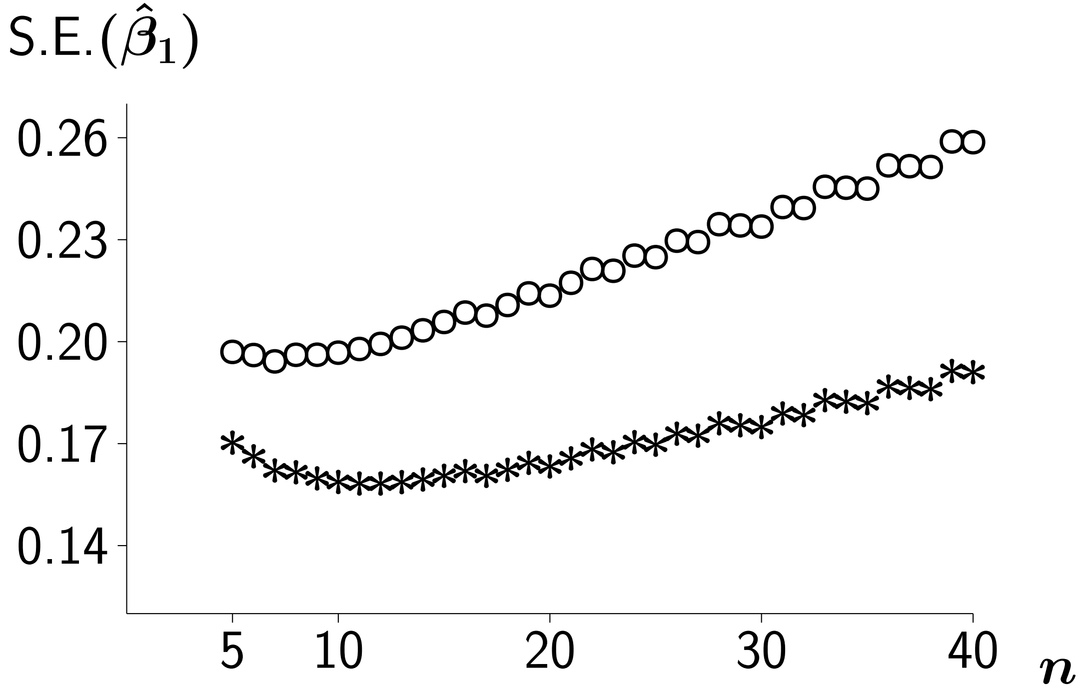


Figure 1 Standard errors for estimating β_1 ,
for $20N + Nn \leq 1{,}000$;
$*$ for $\sigma^2 = 0.6,\ \tau_0^2 = 0.1$;
$\circ$ for $\sigma^2 = 0.5,\ \tau_0^2 = 0.2$.

For these parameters, $\boldsymbol{7 \leq n \leq 17}$ acceptable.

Sensitivity analysis: also calculate for $\boldsymbol{\sigma^2 = 0.5,\ \tau_0^2 = 0.2}$.
(Residual intraclass correlation twice as big.)
Now $\boldsymbol{5 \leq n \leq 12}$ acceptable.

Estimating intraclass correlation

Donner (ISR 1986):

$$\text{S.E.}(\hat{\rho}_I) = (1 - \rho_I) \times (1 + (n-1)\rho_I)\sqrt{\frac{2}{n(n-1)(N-1)}} .$$

Budget constraint:
substitute $N = k/(c_1 + c_2 n)$ and plot as a function of n for various ρ_I.

E.g., suppose $20\,N + Nn \leq 1000$
and $0.1 \leq \rho_I \leq 0.2$.

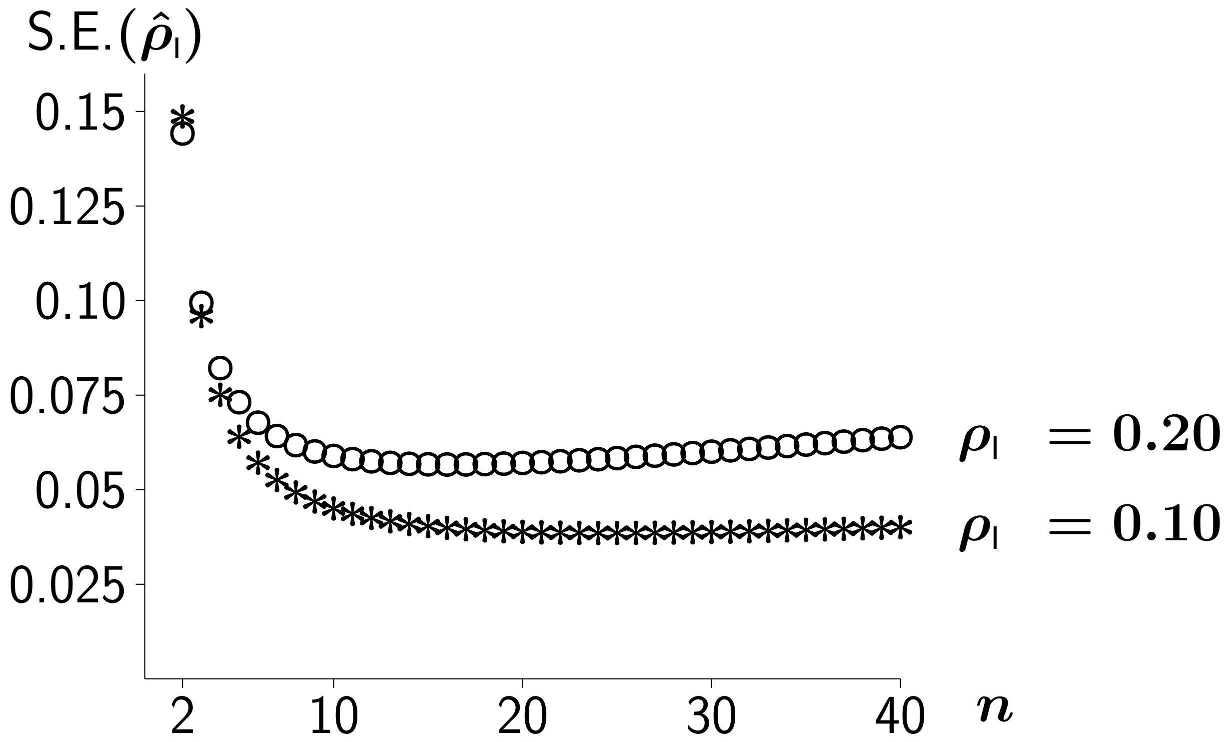


Figure 2 Standard error for estimating intraclass correlation coefficient for budget constraint $20N + Nn \leq 1000$ with $\rho_I = 0.1$ and 0.2.

Clusters sizes between 16 and 27 are acceptable.

Variance parameters

Approximate formulae for random intercept model (Longford, 1993)

$$\text{S.E.}(\hat{\sigma}^2) \approx \sigma^2 \sqrt{\frac{2}{N\,(n-1)}}$$

and

$$\text{S.E.}(\hat{\tau}_0^2) \approx \sigma^2 \, \frac{2}{N\,n} \sqrt{\frac{1}{n-1} + \frac{2\,\tau_0^2}{\sigma^2} + \frac{n\,\tau_0^4}{\sigma^4}} \, .$$

Standard errors for estimated standard deviations:

$$\text{S.E.}(\hat{\sigma}) \approx \frac{\text{S.E.}(\hat{\sigma}^2)}{2\,\sigma} \, ,$$

and similarly for S.E.$(\hat{\tau}_0)$.

Same procedure:
substitute $N = k/(c_1 + c_2 n)$ and plot as function of n.
Example in Cohen (J. Off. Stat., 1998).

12.2 Sandwich estimators

Cluster-robust standard errors for fixed effects can be used to account for clustering (a nested dependence structure) in the data, even when the fixed effects are estimated by a model without the proper dependence assumptions.
(This estimation method is called GEE = Generalized Estimating Equations.)

These are also affectionately called *sandwich estimator* because of their sandwich form

$$\text{var}(\hat{\gamma}) = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} .$$

Here $\boldsymbol{X}$ is the design matrix for the fixed effects, $\boldsymbol{W}$ is the inverse of the assumed covariance matrix in the clusters, and $\boldsymbol{V}$ is an estimate of the within-cluster residual variances.

This way of calculating standard errors does not rely on a particular random effects specification, nor on normality of the residuals.
This works well for 30 or more groups if the groups are similar in size and in the structure of the covariates.

The advantage is robustness; the disadvantage is non-optimal efficiency.

Another issue is that the model is incompletely specified: the random part of the model is ignored.

Considerations about whether or not to do this:

1. Can the research question be answered with the incompletely specified model? (Cf. the nuisance-interest contrast.)
   Are only the fixed parameters of interest?
2. Is the model specification of the random part adequate?
3. Are sample sizes etc. sufficient to make the sandwich estimator a reliable estimator of variance?
4. Is the loss of efficiency acceptable?
   For the latter two questions, the degree of imbalance between the clusters will be an issue.

13. Imperfect Hierarchies

So far we had perfect nesting of lower in higher level units

Non-nested structures:

1. Cross-classified
2. Multiple membership
3. Multiple membership & multiple classification

Cross-classification

Individuals are simultaneously members of several social units (such as neighborhood and schools)

This leads to crossed factors (random effects)

Individuals uniquely belong to a combination of both factors

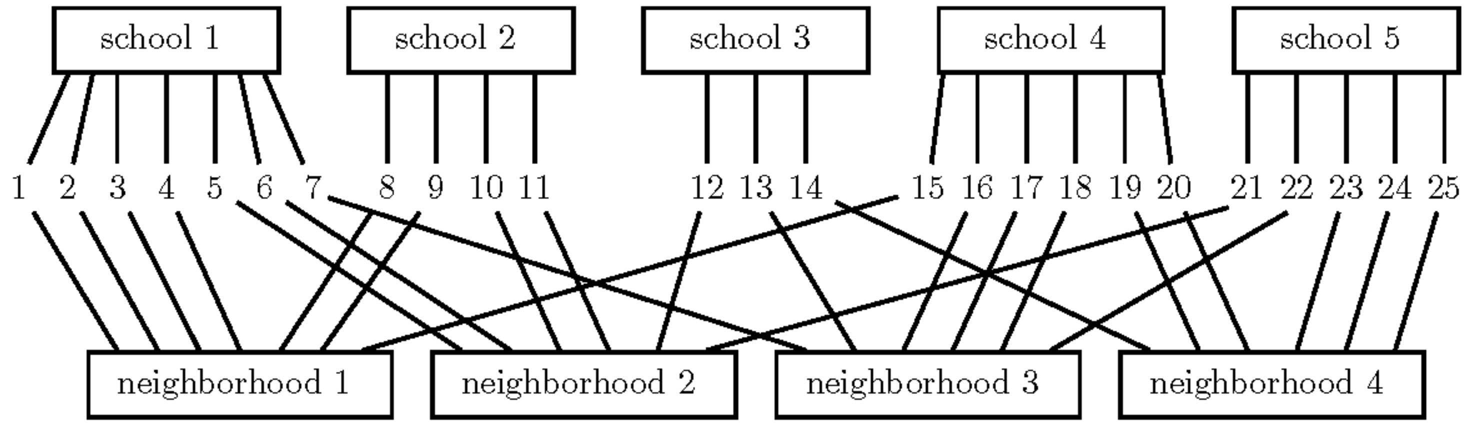


Figure 13.1 Example of pupils nested in schools and neighborhoods

Adapted notation

Level 1 unit (say, pupils) indicated by $\boldsymbol{i}$, simultaneously nested in

- Level 2a (schools) indexed by $\boldsymbol{j}$ and
- Level 2b (neighborhoods) indexed by $\boldsymbol{k}$

Indicate the neighborhood of a pupil $\boldsymbol{i}$ in school $\boldsymbol{j}$ as $\boldsymbol{k(i,j)}$

HLM for a non-nested model:

$$\boldsymbol{Y_{ij} = \gamma_0 + \sum_{h=1}^{q} \gamma_h \, x_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj} \, x_{hij} + R_{ij}}$$

Add neighborhood random effect $\boldsymbol{W_{0k}}$ assuming

$$\boldsymbol{W_{0k} \sim N(0, \tau_W^2)}$$

and

$$\text{Cov}(\boldsymbol{W_{0k}, R_{ij}}) = \boldsymbol{0}, \quad \text{Cov}(\boldsymbol{W_{0k}, U_{0j}}) = \boldsymbol{0}$$

HLM with neighborhood random effect $\boldsymbol{W_{0k}}$:

$$\boldsymbol{Y_{i(j,k)} = \gamma_0 + \sum_{h=1}^{q} \gamma_h\, x_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj}\, x_{hij} + W_{0k} + R_{ij}}$$

Intra-class correlations:

1. Units $\boldsymbol{i}$ within same $\boldsymbol{j}$, different $\boldsymbol{k}$

$$\boldsymbol{\frac{\tau_W^2}{\tau_W^2 + \tau_0^2 + \sigma^2}}$$

2. Units $\boldsymbol{i}$ within same $\boldsymbol{k}$, different $\boldsymbol{j}$

$$\boldsymbol{\frac{\tau_0^2}{\tau_W^2 + \tau_0^2 + \sigma^2}}$$

3. Units $\boldsymbol{i}$ within same $\boldsymbol{j}$, same $\boldsymbol{k}$

$$\boldsymbol{\frac{\tau_W^2 + \tau_0^2}{\tau_W^2 + \tau_0^2 + \sigma^2}}$$

Example: Decomposing complex school effects

Secondary school examination scores of cohort study of Dutch pupils

3,658 pupils in 185 secondary schools coming from 1,292 primary schools

Are primary school effects persistent?

Models:

1. secondary school effects only
2. crossed primary and secondary effects

Results for models with and without cross-classified random effects

	Model 1		Model 2	
Fixed effect	Coeff.	S.E.	Coeff.	S.E.
γ_0 Intercept	6.36	0.03	6.36	0.03
Random part	Var. comp.	S.E.	Var. comp.	S.E.
Crossed random effect:				
$\tau_W^2 = \text{var}(W_{0k})$ primary school			0.006	0.005
Level-two random effect:				
$\tau_0^2 = \text{var}(U_{0j})$ secondary school	0.067	0.014	0.066	0.014
Level-one variance:				
$\sigma^2 = \text{var}(R_{ij})$	0.400	0.010	0.395	0.010

Three intra-class correlations

1. Correlation between examination grades of pupils who attended the same primary school but went to a different secondary school
$$\frac{\tau_W^2}{\tau_W^2 + \tau_0^2 + \sigma^2} = \frac{0.006}{0.467} = 0.013;$$

2. Correlation between grades of pupils who attended the same secondary school but came from different primary schools
$$\frac{\tau_0^2}{\tau_W^2 + \tau_0^2 + \sigma^2} = \frac{0.066}{0.467} = 0.141;$$

3. Correlation between grades of pupils who attended both the same primary and the same secondary school
$$\frac{\tau_W^2 + \tau_0^2}{\tau_W^2 + \tau_0^2 + \sigma^2} = \frac{0.072}{0.467} = 0.154.$$

Model with added fixed effects

	Model 2		Model 3	
Fixed effect	Coeff.	S.E.	Coeff.	S.E.
γ_0 Intercept	6.36	0.03	6.39	0.02
γ_1 Pretest			0.032	0.002
γ_2 SES			0.068	0.011
γ_3 Advice			0.054	0.009
γ_4 Ethnicity			−0.071	0.028
Random part	Var. comp.	S.E.	Var. comp.	S.E.
$\tau_W^2 = \text{var}(W_{0k})$ primary school	0.006	0.005	0.003	0.003
$\tau_0^2 = \text{var}(U_{0j})$ secondary school	0.066	0.014	0.034	0.008
$\sigma^2 = \text{var}(R_{ij})$	0.395	0.010	0.330	0.008

Multiple memberships

Individuals are/have been members of several social units
(such as several different secondary schools)

Only one random factor at level two
(as opposed to two or more factors in cross-classified models)

Individuals belong to multiple levels of that factor

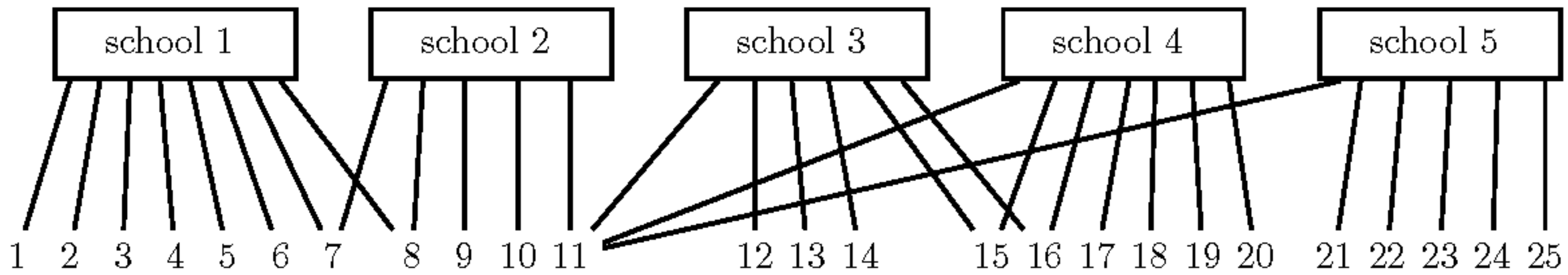


Figure 13.2 Example of pupils nested within multiple schools.

Membership weights

We weight membership by importance
(such as duration of stay in particular school)

Weights w_{ih} for each pupil i in school h, with $\sum_{h=1}^{N} w_{ih} = 1$

Example:

1. Pupil 1: only in school 1
2. Pupil 2: equal time in schools 1 and 3
3. Pupil 3: equal time in schools 1, 2, 3

	School		
Pupil	1	2	3
1	1	0	0
2	0.5	0	0.5
3	0.33	0.33	0.33

Multilevel model for multiple memberships

Denote by $Y_{i\{j\}}$ an individual who might have multiple memberships

Write HLM with level two residuals U_{0h} weighted by w_{ih}

$$Y_{i\{j\}} = \gamma_0 + \sum_{h=1}^{N} w_{ih} U_{0h} + R_{i\{j\}}$$

For example, pupil 2 will have

$$\frac{1}{2} U_{01} + \frac{1}{2} U_{03}$$

To include a fixed effect for membership trajectory define

$$W_i = \frac{1}{\sum_h w_{ih}^2} - 1$$

which is 0 for pupil 1 (with one membership) and positive for others with multiple memberships

Example: Decomposition of school effects with multiple memberships

Continuation of previous example

94% of pupils never changed schools

Remaining 6%

- 215 attend two schools ($\boldsymbol{w}_i$ vary between 0.2 and 0.8)
- 5 attend three ($\boldsymbol{w}_i$ vary between 0.2 and 0.25)

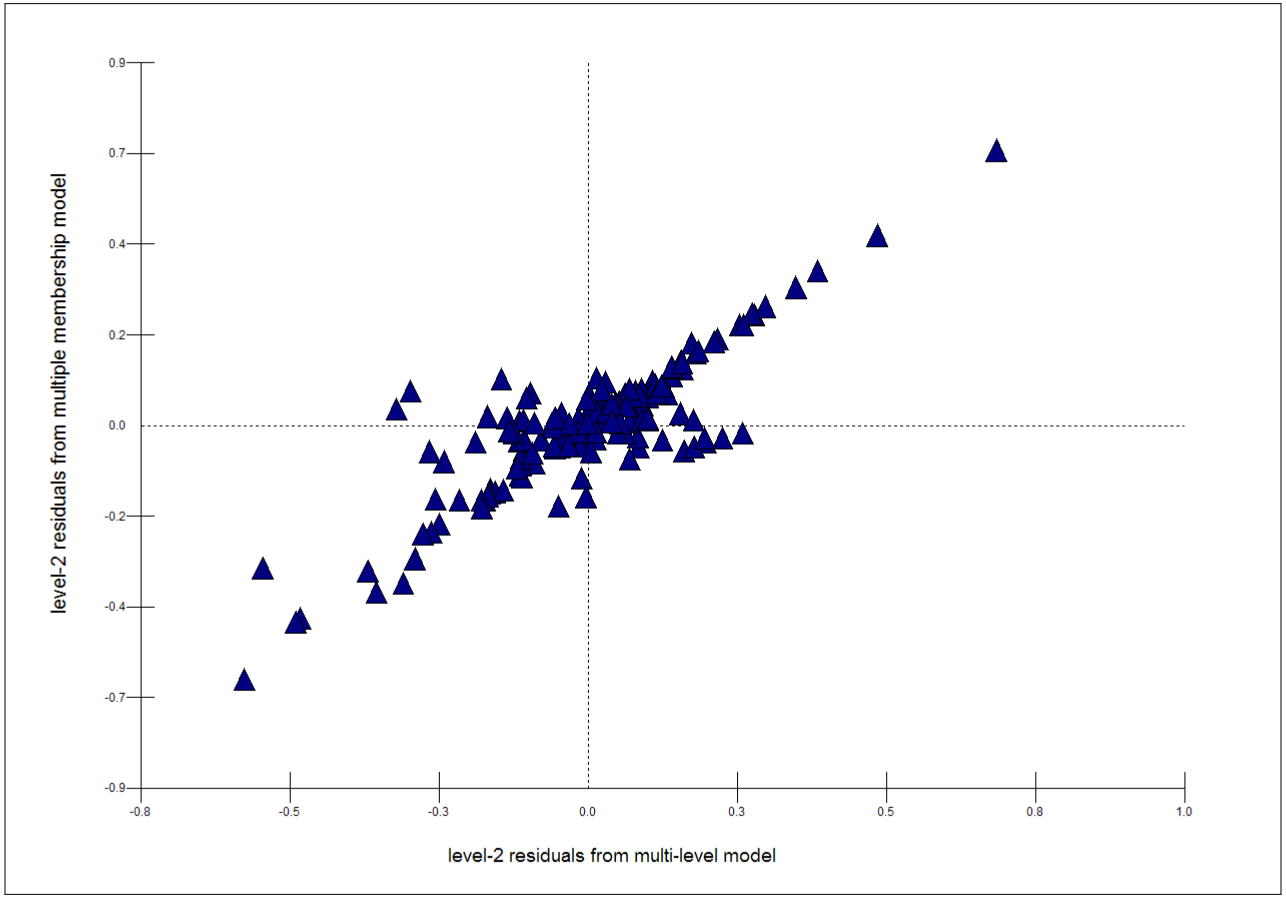
Models:

1. Level 2: only first secondary school
2. Level 2: multiple membership model

Results from models without and with multiple membership

	Model 4 (ML)		Model 5 (MM)	
Fixed effect	Coeff.	S.E.	Coeff.	S.E.
γ_0 Intercept	6.36	0.03	6.36	0.03
Random part	Var. comp.	S.E.	Var. comp.	S.E.
Level-two random effect:				
$\tau_0^2 = \text{var}(U_{0j})$ secondary school	0.062	0.013	0.064	0.014
Level-one variance:				
$\sigma^2 = \text{var}(R_{ij})$	0.402	0.010	0.401	0.009

Plot of school level residuals from ML and MM



Note: Correlation is 0.86

Multiple membership multiple classification models

Combining cross-classified random effects with multiple membership weights

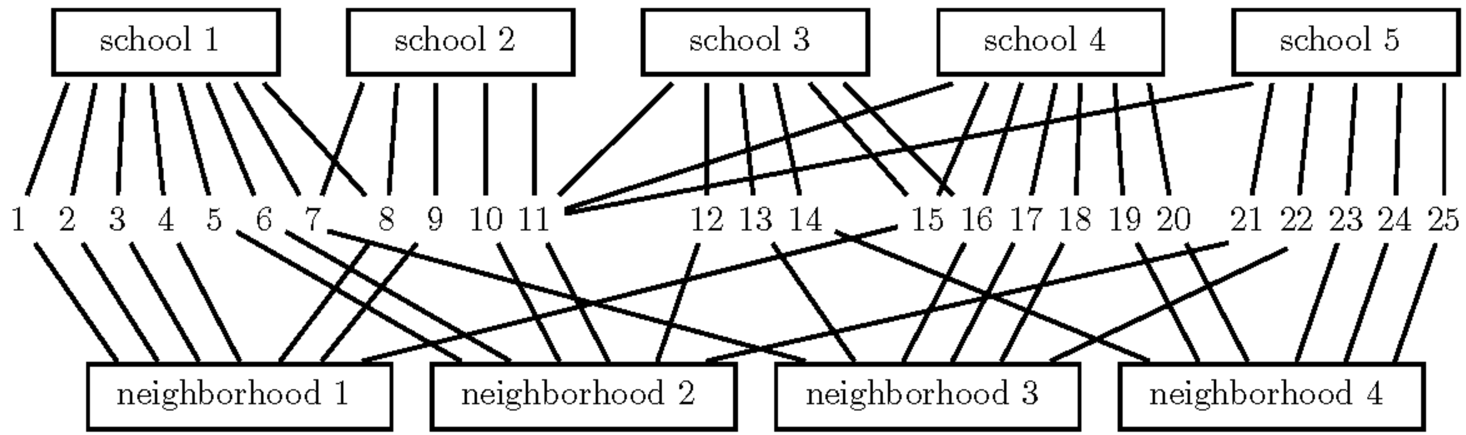


Figure 13.4 Example of pupils nested within multiple schools crossed with neighborhoods

	Model 7		Model 8	
Fixed effect	Coeff.	S.E.	Coeff.	S.E.
γ_0 Intercept	6.36	0.03	6.40	0.02
γ_1 Pretest			0.031	0.002
γ_2 SES			0.068	0.011
γ_3 Advice			0.053	0.010
γ_4 Ethnicity			–0.073	0.028
Random part	Var. comp.	S.E.	Var. comp.	S.E.
Crossed random effect:				
$\tau_W^2 = \text{var}(W_{0k})$ primary school	0.006	0.005	0.003	0.003
Level-two random MM effect:				
$\tau_0^2 = \text{var}(U_{0j})$ secondary school	0.065	0.014	0.033	0.007
Level-one variance:				
$\sigma^2 = \text{var}(R_{ij})$	0.395	0.010	0.331	0.008

14. Survey weights

Surveys often are organized using a *design* in which the sample is not totally random: e.g., a stratified and/or multi-stage sample.

A major issue is the possibility that at some or all levels, units are represented with non-constant probabilities. Reasons can be the efficient organization of data collection, or the wish to over-represent some interesting subgroups.

This is expressed by *sampling weights* (proportional to sampling probabilities).

Non-constant sampling weights imply that the sample cannot be regarded as a 'mini-population' – it is not directly representative.

What to do with these weights?

What is the purpose of the statistical inference:

1. *Descriptive*
   e.g., estimate mean or proportion in a specific population.
2. *Analytic*
   how do variables of interest depend on other variables?
   Here often there is a wish to generalize to a larger population.

Connected to this, but a distinct issue is the fact that the use of a *probability model* can be founded on different bases:

1. *Design-based*
   This uses the probability distribution that is implied by the sampling design.
   This usually is for a sample from a finite population.
   The probabilities are under control of the researcher,
   except for non-response and other sampling errors.
2. *Model-based*
   The researcher assumes that data can be regarded
   as outcomes of a probability model with some unknown parameters.
   This usually assumes a hypothetical infinite population,
   e.g., with normal distributions.
   Assumptions are made about independence etc.

Most of quantitative social science research is about analytic inference.

If the model is true (or a good approximation),
and a survey design is used for which
the sample is drawn independently of the residuals in the statistical model,
then the design is irrelevant and taking it into account leads to a loss of efficiency.

But one can hardly ever (or never) be certain that a model is true
and design-based estimators can give protection against
distortions caused by incorrect model assumptions.

If the design is potentially relevant,
then the contrast nuisance $\Leftrightarrow$ interesting phenomenon occurs again:
the design may tell is something interesting about the social world.

The approach followed here is that we favor a model-based approach,
taking account of the design when this is called for, or when this is interesting.

Two kinds of weights

In statistical modeling, two kind of weights are used.

1. *Sampling weights*, used in complex surveys, which are the inverses of the probabilities of including an element in the sample;
   i.e., population elements that are undersampled
   get a higher weight to correct for the undersampling.
   Purpose: unbiased estimation, correct for non-uniform sampling.

2. *Precision weights*, expressing that some data points are associated with more precision than others, and therefore get a stronger effect on the results.
   Example: multiplicity weights (representing number of cases).
   Purpose: lower estimation variance, i.e., increased precision.

These kinds of weight differ in their implications for the standard errors.

A survey sample with a few cases with very high weights can lead to high standard errors, because the results are dominated by these high-weight cases.

The *effective sample size*, for a sample with weights $\boldsymbol{w_i}$, is

$$\boldsymbol{n}_{\text{eff}} = \frac{\left(\sum_i \boldsymbol{w_i}\right)^2}{\sum_i \boldsymbol{w_i^2}}$$

(applicable for estimating a mean, not for regression).

This is equal to the number of sample elements if $\boldsymbol{w_i}$ is constant,
and else it is smaller: *loss of efficiency*.

The *design effect* is the ratio of effective sample size
to number of sample elements.

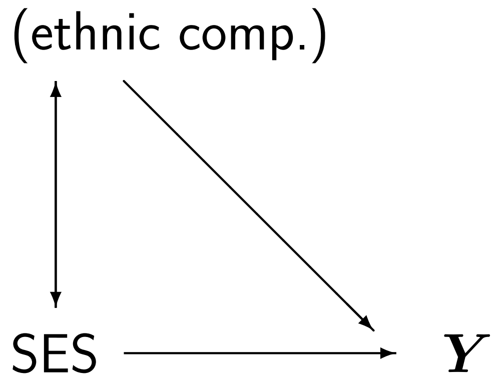
When can sampling weights be ignored?

A sample design will depend on a number of *design variables*, used for stratification, to define primary sampling units, etc. The survey weights are a function of the design variables.

If the model is specified correctly given all the design variables , i.e., the residuals in the model are independent of the design variables, then the sample design can be ignored in the analysis.

Therefore, the sample design is important for analytic inference mainly if we are working with a misspecified model.

As an example, consider a study of pupils in schools, $\boldsymbol{Y}$ = language achievement, with SES as an explanatory variable, and ethnic composition as an omitted variable.



Suppose that the inclusion probability of schools is associated with the ethnic composition, but no data about this are available.

Then the analysis effectively uses a SES variable that also represents the unobserved ethnic composition;
if the analysis is unweighted and sampling is not uniform,
the regression coefficient of SES will average over ethn. comp.
in proportions according to the sample, differring from those in the population.

We argue for aiming, when possible, for a model-based analysis, with the analysis being informed by the design variables.

Arguments:

1. The focus of analysis should be on getting a good model.
2. Most social science researchers are more acquainted with model-based inference, and have less risk of making mistakes there than in design-based inference.
3. Methods for design-based inference for multilevel data structures still have strong limitations.

Thus, the design is used to try and make the model more interesting and appropriate; a design-based analysis is followed only if the model-based analysis seems unreliable.

How to use design variables in a model-based analysis?

1. Check the variability of weights at each level.

If weights are almost constant, then they may be ignored;
if weights are strongly variable, then using them is risky.

2. Add design variables to the model of interest.

If possible, this is the golden way.
Design variables should be considered also for the random part and in interactions.
Design variables are, however, not always available for the researcher.

3. Apply the hierarchical linear model separately to different parts of the survey.

These parts may depend on stratification variables, on design weights, etc.

Grouping according to design weights is important here,
because the assumption for a model-based analysis is that
the model does not depend on the design weights.

4. Add weight variables to the model of interest.

This is an option if the design variables themselves are not available.

If the weight variables themselves do not have a clear interpretation, the purpose of this is mainly the hope that including the weight variables does not affect the results for the other variables, which would be an argument for pursuing a model-based analysis without the weight variables.

5. Carry out model-based and design-based elements within each level-2 unit, and compare the results.

For a two-level design, this leads to single-level within-group analyses. The differences between the two kinds of analyses can be inspected and tested.

Again, this is done with the hope of being able to choose for a model-based approach; if this is not justified, then inspection of the differences can give insight into how the design is associated with the variables of interest.

6. Test differences between the model-based and design based results.

Equation (4.17) gives Asparouhov's (2006) measure for the informativeness of the sample design, based on such a comparison:

$$\boldsymbol{I_2} = \frac{\hat{\gamma}_{\boldsymbol{h}}^{\mathbf{HLM}} - \hat{\gamma}_{\boldsymbol{h}}^{\boldsymbol{W}}}{\text{S.E.}\left(\hat{\gamma}_{\boldsymbol{h}}^{\mathbf{HLM}}\right)} ,$$

the difference between the estimates,
expressed in units relative to the model-based standard error.

Example: Metacognitive strategies in PISA 2009 for USA

PISA is a large international OECD study about school achievement.

We consider the PISA 2009 data for the USA,
with the dependent variable 'METASUM', a measure of awareness of appropriate strategies for summarizing information.

Sampling design: stratified two-stage sample
Stratification by school type {private, public} × region {MW, NE, S, W}.
Schools are sampled within strata with probabilities proportional to enrollment.

Numbers of sampled schools (left) and sampled pupils (right) per stratum, for PISA data, USA, 2009.

	Midwest		Northeast		South		West	
Public schools	38	1,264	26	732	55	1,776	35	1,116
Private schools	2	71	2	46	4	108	3	120

There are too few private schools to say much specifically about them.

Level-one weights all are between .95 and 1.00: can be disregarded.

Level-two design effects for PISA data (USA 2009).

	Midwest	Northeast	South	West
Public schools	0.49	0.74	0.16	0.17

For South and West, this is reason for concern.

Example of a descriptive analysis

Suppose we wish to estimate the proportions of schools in the South located in villages, towns, and cities.

Note that villages tend to have smaller, and cities larger schools.

Design-based and model-based estimates for three proportions,
for public schools in the South.

Parameter	Probability-weighted		Unweighted	
Proportion of schools situated in a	$\hat{\beta}^p$	S.E.^{p}	$\bar{y}$	S.E.$(\bar{y})$
Village	0.18	0.09	0.11	0.04
Town	0.65	0.13	0.56	0.07
City	0.17	0.07	0.33	0.06

The unweighted results are seriously biased for villages and cities,
and unweighted standard errors are unreliable.

Analysis of metacognitive competences

Suppose we wish to study how metacognitive competences depend on the following variables.

- Gender; female = 0, male = 1.
- Grade, originally ranging from 8 to 12; centered at 10, new range –2 to +2.
- Age, between 15 and 17 years, centered at 16 years.
- ESCS, the PISA index of economic, social and cultural status of students; in this data set this has mean 0.15 and standard deviation 0.92.
- Immigration status, recoded to:
  0, at least one parent born in the USA;
  1, second generation (born in the USA, both parents born elsewhere);
  2: first generation (born outside the USA, parents likewise).
  This ordinal variable is treated as having a linear effect.
- Public versus private schools.

To explore the consequences of the design,
first we analyze the data set in five separate parts:

private schools;
schools in the four quartiles of the level-two weights.

Estimates for model for metacognitive competence for five parts of the data set: fixed effects.

	Private		Weight 1		Weight 2		Weight 3		Weight 4	
N(schools)	11		38		39		38		39	
Fixed effects	Par.	S.E.	Par.	S.E.	Par.	S.E.	Par.	S.E.	Par.	S.E.
Intercept	–0.72	0.48	–0.12	0.10	–0.15	0.10	–0.17	0.06	–0.18	0.07
Male	–0.27	0.12	–0.34	0.06	–0.33	0.06	–0.28	0.06	–0.35	0.06
Age	–0.05	0.23	–0.05	0.12	–0.18	0.12	–0.26	0.11	–0.01	0.12
Grade	0.25	0.12	0.20	0.06	0.24	0.07	0.28	0.06	0.18	0.07
Immigrant	–0.13	0.09	0.01	0.04	0.05	0.05	–0.03	0.06	0.06	0.09
ESCS	0.04	0.08	0.10	0.04	0.14	0.04	0.08	0.04	0.11	0.04
Sch-imm	0.51	0.37	0.04	0.14	0.14	0.16	0.10	0.16	–0.29	0.36
Sch-ESCS	0.74	0.40	0.16	0.12	0.11	0.12	0.25	0.13	0.10	0.13

Estimates for model for metacognitive competence for five parts of the data set: variance parameters.

	Private	Weight 1	Weight 2	Weight 3	Weight 4
N (schools)	11	38	39	38	39
Variances	Var.	Var.	Var.	Var.	Var.
School lev.	0.14	0.04	0.06	0.02	0.05
Student lev.	0.85	0.93	0.97	0.95	0.95

Conclusions:

Private schools differ strongly from public schools;
the four weight groups differ w.r.t. school-level variables.

*Which other school variables are available
that might explain these differences?*

A major difference between the weight groups is urbanization.

Next page:

Estimates for model for metacognitive competence, including urbanization, for five parts of the data set: fixed effects.

Reference category for urbanization is 'city'.

	Private		Weight 1		Weight 2		Weight 3		Weight 4	
N schools	11		38		39		38		39	
Fixed effects	Par.	S.E.	Par.	S.E.	Par.	S.E.	Par.	S.E.	Par.	S.E.
Intercept	–0.90	0.40	–0.16	0.12	0.03	0.13	0.01	0.13	–0.01	0.23
Male	–0.27	0.11	–0.33	0.06	–0.33	0.06	–0.28	0.06	–0.35	0.06
Age	–0.07	0.23	–0.04	0.12	–0.19	0.12	–0.27	0.11	–0.01	0.12
Grade	0.26	0.12	0.20	0.06	0.25	0.07	0.28	0.06	0.18	0.07
Immigrant	–0.13	0.09	0.01	0.04	0.05	0.05	–0.03	0.06	0.06	0.09
ESCS	0.03	0.08	0.10	0.04	0.14	0.04	0.08	0.04	0.11	0.04
Sch-imm	0.85	0.33	0.11	0.15	–0.00	0.18	0.11	0.17	0.12	0.43
Sch-ESCS	1.06	0.33	0.19	0.13	0.14	0.13	0.21	0.14	0.15	0.14
Large city	–0.80	0.26	–0.10	0.12	0.01	0.17	–0.32	0.18	–0.51	0.34
Town	–0.36	0.24	0.12	0.12	–0.17	0.13	–0.20	0.13	–0.38	0.26
Small town	—		0.02	0.21	–0.42	0.17	–0.22	0.13	–0.24	0.25
Village	—		—		—		–0.23	0.17	–0.12	0.24

Estimates for model for metacognitive competence, including urbanization, for five parts of the data set: variance parameters.

	Private	Weight 1	Weight 2	Weight 3	Weight 4
N schools	11	38	39	38	39
Variances	Var.	Var.	Var.	Var.	Var.
School lev.	0.05	0.04	0.05	0.02	0.05
Student lev.	0.85	0.93	0.97	0.95	0.95

Conclusion

Differences between weights groups are very small now.

Let us at this stage compare the model-based and design-based estimates.

Next page:

Design-based and model-based estimates for model for metacognitive competence, entire data set: fixed effects.

$\boldsymbol{I_2}$ is Asparouhov's (2006) informativeness measure.

	Design-based		Model-based		
Fixed effects	Par.	S.E.	Par.	S.E.	I_2
Intercept	–0.077	0.089	–0.075	0.058	0.03
Male	–0.350	0.040	–0.321	0.029	1.00
Age	–0.134	0.102	–0.121	0.055	0.24
Grade	0.192	0.043	0.224	0.031	1.03
Immigrant	–0.014	0.031	0.008	0.027	0.81
ESCS	0.075	0.035	0.107	0.019	1.68
Sch-imm	0.240	0.104	0.094	0.068	2.41
Sch-ESCS	0.426	0.119	0.189	0.049	4.84
Private	–0.211	0.179	–0.008	0.118	1.72
Large city	–0.406	0.121	–0.152	0.081	3.14
Town	–0.390	0.111	0.093	0.057	8.47
Small town	–0.382	0.108	0.172	0.072	7.69
Village	–0.203	0.110	–0.087	0.073	1.59

Design-based and model-based estimates for model for metacognitive competence, entire data set: variance parameters.

	Design-based		Model-based	
Variances	Var.		Var.	
School lev.	0.031	0.007	0.040	0.009
Student lev.	0.893	0.083	0.941	0.018

Conclusion:

Differences remain, mainly with respect to school-level variables.

A residual analysis showed that there is one outlier:
a school with 6,694 pupils enrolled,
while other schools range between 100 and 3,592.

This school and all private schools were excluded,
and square root of school size included as control variable.

Next page:

Design-based and model-based estimates for model for metacognitive competence, public schools without outlier: fixed effects.

$\boldsymbol{I_2}$ is Asparouhov's (2006) informativeness measure.

	Design-based		Model-based		
Fixed effects	Par.	S.E.	Par.	S.E.	I_2
Intercept	–0.050	0.068	–0.122	0.061	1.18
Male	–0.359	0.046	–0.318	0.032	1.28
Age	0.011	0.080	–0.117	0.059	2.17
Grade	0.162	0.040	0.215	0.032	1.66
Immigrant	0.019	0.034	0.017	0.030	0.07
ESCS	0.090	0.032	0.111	0.019	1.11
Sch-imm	0.114	0.092	0.068	0.078	0.59
Sch-ESCS	0.204	0.081	0.143	0.054	1.13
$\sqrt{\text{school size}} - 35$	0.004	0.005	0.004	0.003	0.00
Large city	–0.193	0.091	–0.094	0.078	1.27
Town	–0.094	0.059	–0.032	0.060	1.03
Small town	–0.128	0.088	–0.100	0.080	0.35
Village	0.034	0.102	0.023	0.088	0.13

Design-based and model-based estimates for model for metacognitive competence, public schools without outlier: variance parameters.

	Design-based		Model-based	
Variances	Var.		Var.	
School lev.	0.035	0.010	0.038	0.009
Student lev.	0.939	0.085	0.940	0.018

Conclusion:

Only the coefficient of age still differs between the two types of analysis.
This suggests that age effects depend on the design variables.
This could be, e.g., school enrollment or urbanization.

Next page:

Design-based and model-based estimates for model for metacognitive competence, public schools without outlier, with more extensive controls: fixed effects.

$\boldsymbol{I_2}$ is Asparouhov's (2006) informativeness measure.

	Design-based		Model-based		
Fixed effects	Par.	S.E.	Par.	S.E.	I_2
Intercept	–0.067	0.065	–0.120	0.061	0.87
Male	–0.357	0.045	–0.319	0.032	1.19
Age	–0.083	0.065	–0.113	0.058	0.52
Grade	0.167	0.040	0.216	0.032	1.53
Immigrant	0.019	0.034	0.017	0.030	0.07
ESCS	0.090	0.032	0.111	0.019	1.11
Sch-imm	0.108	0.092	0.065	0.060	0.72
Sch-ESCS	0.203	0.081	0.143	0.054	1.11
$\sqrt{\text{school size}} - 35$	0.002	0.005	0.003	0.003	0.33
Large city	–0.194	0.091	–0.093	0.078	1.29
Town	–0.096	0.060	–0.033	0.060	1.05
Small town	–0.134	0.089	–0.101	0.080	0.41
Village	0.036	0.104	0.023	0.088	0.15
Age $\times$ ($\sqrt{\text{school size}} - 35$)	–0.011	0.005	–0.004	0.004	1.75

Design-based and model-based estimates for model for metacognitive competence, public schools without outlier, with more extensive controls: variance parameters.

	Design-based		Model-based	
Variances	Var.		Var.	
School lev.	0.036	0.010	0.038	0.009
Student lev.	0.937	0.084	0.940	0.018

Conclusion:

It seems that now we arrived at a good model.

Overview: how was the analysis modified by the design variables

1. Public schools differ from private schools,
   but the data contains too few private schools to say much about them.
2. One very large school seemed to differ from the rest.
3. Urbanization has an effect on metacognitive competences.
4. Age (controlling for grade!) has an effect,
   and interacts with urbanization and/or school size;
   it is hard to say which of these, because of correlated effects.
5. For the student-level variables, the final model-based results differ hardly from the model-based results for the entire data set, controlling for urbanization.
   There is a difference w.r.t. the school-level variable 'proportion of immigrants', which may be due to the difference between private and public schools.
   Here again there is an issue of correlated effects.
6. The refinement of excluding the very large school and using school size as a control variable did not importantly modify the other results.

15. Longitudinal data

Level two: *'subjects'*;
level one: *'measurements'* made at certain time points.

Multilevel ('random coefficients') approach: flexible model specification, easy treatment of unbalanced data sets (variable n per subject).

1. *Fixed occasions* :
   All subjects measured at the same time points.
2. *Variable occasions* :
   time points can be different; also number of measurements may differ.
   (Then there are few alternative approaches.)

An advantage in both cases is the interpretability of the parameters in terms of rate of increase, etc.

Notation: Y_{ti} is measurement at time t for individual i.
t indicates level 1, i indicates level 2.

15.1 Fixed Occasions

Simplest case: random intercept model.
This is also called the 'compound symmetry' model.

$$Y_{ti} \; = \; \mu_t \, (+ \text{ further fixed part}) \; + \; U_{0i} \; + \; R_{ti}.$$

The subjects differ only with respect to their constant mean deviations U_{0i},
the time points could have arbitrary means μ_t .
The residuals are independent.

The intraclass correlation here compares individual differences to over-time fluctuations, and will be much larger than the .05–.25 range that is usual for data sets of individuals nested in groups.

Arbitrary differences between time points can be represented by defining dummy variables for all measurement occasions:

$$\sum_h \mu_h d_{hti} \; + \; U_{0i} \; + \; R_{ti} \text{ where } d_{hti} = 1 \text{ for } h = t, 0 \text{ otherwise}$$

(or all minus one, if a constant term — intercept — is included).

Example: Life Satisfaction of German 55-60 year olds

1236 respondents, panel data with yearly waves;
Fixed occasions: $\boldsymbol{t = 1, 2, 3, 4, 5, 6}$ representing ages $\mathbf{55, 56, \ldots, 60}$.

The first model has a constant mean over time,
the second an age-dependent mean.

Table 12.1 Estimates for random intercept models

		Model 1		Model 2	
Fixed effect		Coefficient	S.E.	Coefficient	S.E.
μ_1	Mean at age 55	6.937	0.044	6.882	0.053
μ_2	Mean at age 56	6.937	0.044	6.956	0.054
μ_3	Mean at age 57	6.937	0.044	7.021	0.056
μ_4	Mean at age 58	6.937	0.044	6.907	0.057
μ_5	Mean at age 59	6.937	0.044	6.894	0.059
μ_6	Mean at age 60	6.937	0.044	6.985	0.060
Random effect		Parameter	S.E.	Parameter	S.E.
Level-two (i.e., individual) variance:					
$\tau_0^2 = \text{var}(U_{0i})$		1.994	0.095	1.991	0.094
Level-one (i.e., occasion) variance:					
$\sigma^2 = \text{var}(R_{ti})$		1.455	0.030	1.452	0.030
Deviance		21,791.34		21,780.70	

Model 2 is borderline significantly better than Model 2
($\chi^2 = 10.64,\ df = 5, p = 0.06$),
but the mean differences are not important.

Estimated variances of observations are $\hat{\tau}_0^2 + \hat{\sigma}^2 = 1.994 + 1.455 = 3.449$,
estimated within-subjects correlation is

$$\hat{\rho}_I = \frac{1.994}{1.994 + 1.455} = 0.58 \ .$$

Note that these correlations are the same between all measurements of the same subject, whether they are for adjacent waves or for waves spaced far apart.

If we wish to estimate a person's long-term life satisfaction $\mu + U_{0i}$,
then the reliability of measurement when using n repeated measurements is

$$\lambda_i = \frac{n\rho_{\mathrm{I}}}{1 + (n-1)\rho_{\mathrm{I}}}$$

(see Sections 3.5 and 4.8); e.g., for $n = 3$ this is 0.81.
For this purpose again we can use the posterior means.

Here again the assumptions of the random intercept model are quite restrictive.
It is likely that individual respondents differ not only in their mean value over time, but also in rate of change and other aspects of time dependence.

This is modeled by including random slopes of time,
and of non-linear transformations of time.

Random slope of time:

$$\boldsymbol{Y_{ti} = \mu_t + U_{0i} + U_{1i}\,(t - t_0) + R_{ti}}\ ;$$

with a covariate, e.g.,

$$\boldsymbol{Y_{ti} = \mu_t + \alpha z_i + \gamma z_i(t - t_0) + U_{0i} + U_{1i}\,(t - t_0) + R_{ti}}\ .$$

Here we use as covariates:
birth year (representing cohort & time) and interaction birth year $\times$ age.
Birth year date ranges 1929-1951, coded here as

$$\frac{\text{birth year date} - \mathbf{1940}}{\mathbf{10}}$$

so that it ranges between –1.1 and $+1.1$;
division by 10 to avoid very small coefficients.

Note that age is included non-linearly as a main effect (dummy variables) and linearly in interaction with birth year.

Linear age $(\boldsymbol{t} - \boldsymbol{t}_0)$ ranges from 0 to 5.

The choice of variables implies that the intercept refers to individuals born in 1940 and now aged 55 years.

	Model 3		Model 4	
Fixed effect	Coefficient	S.E.	Coefficient	S.E.
μ_1 Effect of age 55	6.883	0.055	6.842	0.055
μ_2 Effect of age 56	6.955	0.055	6.914	0.055
μ_3 Effect of age 57	7.022	0.055	6.988	0.055
μ_4 Effect of age 58	6.912	0.056	6.886	0.057
μ_5 Effect of age 59	6.906	0.058	6.892	0.060
μ_6 Effect of age 60	7.004	0.060	7.005	0.064
α Main effect birth year			–0.394	0.078
γ Interaction birth year × age			0.049	0.019
Random effect	Parameter	S.E.	Parameter	S.E.
Level-two variation:				
τ_0^2 Intercept variance	2.311	0.126	2.253	0.124
τ_1^2 Slope variance age	0.025	0.005	0.025	0.005
τ_{01} Intercept–slope covariance	–0.103	0.021	–0.097	0.021
Level-one (i.e., occasion) variance:				
σ^2 Residual variance	1.372	0.031	1.371	0.031
Deviance	21,748.03		21,722.45	

Interpretation :

Random slope is significant: $\chi^2 = 32.7,\ df\ = 2,\ p < 0.0001$.

Birth year has effect –0.394: older cohorts tend to be happier;
its contribution for $-1.1 \leq$ B.Y. ≤ 1.1 ranges from 0.43 to –0.43.

This is of medium effect size given the inter-individual s.d.$(U_{0i}) = \hat{\tau}^0 = 1.5$.

The interaction birth year $\times$ age has a somewhat smaller contribution,
ranging between $\pm(5 \times 1.1 \times 0.049) = \pm 0.27$.

Deviations from average pattern in rate of change
have standard error $\sqrt{0.025} = 0.16$.

Since age ranges from 0 to 5, this represents larger variations in rate of change
than those accounted for by birth year.

Negative intercept-slope covariance:
those who start higher (at 55 years) tend to go down
compared to those who start lower.

The covariance matrix implied by this model is

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{Y}^{c}) = \begin{pmatrix} \mathbf{3.67} & \mathbf{2.16} & \mathbf{2.06} & \mathbf{1.96} & \mathbf{1.87} & \mathbf{1.77} \\ \mathbf{2.16} & \mathbf{3.50} & \mathbf{2.01} & \mathbf{1.94} & \mathbf{1.87} & \mathbf{1.80} \\ \mathbf{2.06} & \mathbf{2.01} & \mathbf{3.38} & \mathbf{1.92} & \mathbf{1.87} & \mathbf{1.82} \\ \mathbf{1.96} & \mathbf{1.94} & \mathbf{1.92} & \mathbf{3.31} & \mathbf{1.87} & \mathbf{1.85} \\ \mathbf{1.87} & \mathbf{1.87} & \mathbf{1.87} & \mathbf{1.87} & \mathbf{3.29} & \mathbf{1.88} \\ \mathbf{1.77} & \mathbf{1.80} & \mathbf{1.82} & \mathbf{1.85} & \mathbf{1.88} & \mathbf{3.33} \end{pmatrix}$$

and the correlation matrix is

$$\hat{\boldsymbol{R}}(\boldsymbol{Y}^{c}) = \begin{pmatrix} \mathbf{1.00} & \mathbf{0.60} & \mathbf{0.58} & \mathbf{0.56} & \mathbf{0.54} & \mathbf{0.51} \\ \mathbf{0.60} & \mathbf{1.00} & \mathbf{0.58} & \mathbf{0.57} & \mathbf{0.55} & \mathbf{0.53} \\ \mathbf{0.58} & \mathbf{0.58} & \mathbf{1.00} & \mathbf{0.57} & \mathbf{0.56} & \mathbf{0.54} \\ \mathbf{0.56} & \mathbf{0.57} & \mathbf{0.57} & \mathbf{1.00} & \mathbf{0.57} & \mathbf{0.56} \\ \mathbf{0.54} & \mathbf{0.55} & \mathbf{0.56} & \mathbf{0.57} & \mathbf{1.00} & \mathbf{0.57} \\ \mathbf{0.51} & \mathbf{0.53} & \mathbf{0.54} & \mathbf{0.56} & \mathbf{0.57} & \mathbf{1.00} \end{pmatrix}.$$

The variance does not change a lot with age,
and correlations attenuate slightly as age differences increase
(in the matrix, this means going away from the main diagonal);
they are close to the intra-subject correlation
estimated for the compound symmetry model as $\hat{\rho}_I = 0.58$.

Multivariate model: unstructured general covariance matrix.

A further possibility available for fixed occasion data
(not in this way for variable occasion data) is the fully multivariate model, which makes no restrictions on the covariance matrix.

The advantage is that there is less concern about failure of the assumptions.

In some software implementations (e.g., MLwiN), this can be achieved
by giving all dummy variables for measurement occasions random slopes at level 2, and dropping the random part at level 1.

Simplest case: *incomplete bivariate data*.

This is like the combination of a paired-samples t-test
and an independent-samples t-test.

For the REML estimation method, this method reproduces
the paired-samples t-test if data are complete
and the independent-samples t-test if no respondent has both data points;
but still allowing for different variances at both time points.

The model fitted is

$$Y_{ti} = \gamma_0 + \gamma_1 d_{1ti} + U_{ti},$$

where the dummy variable d_1 indicates whether or not $t = 1$

Example: comparison of life satisfaction between ages 50 and 60,
for individuals born in 1930-1935 and alive in 1984.

Available respondents: 591; of these,
136 with measurements for both ages,
112 with a measurement only for age 50 ($t = 0$),
343 with a measurement only for age 60 ($t = 1$)..

Null hypothesis of no age effects: '$\gamma_1 = 0$'.

Estimates for incomplete paired data.

Fixed effect	Coefficient	S.E.
γ_0 Constant term	7.132	0.125
γ_1 Effect time 1	–0.053	0.141
Deviance	2,953	

Estimated covariance matrix

$$\widehat{\boldsymbol{\Sigma}}(\boldsymbol{Y}^{c}) = \begin{pmatrix} \mathbf{4.039} & \mathbf{1.062} \\ \mathbf{1.062} & \mathbf{3.210} \end{pmatrix}.$$

Age has no significant effect: $\hat{\gamma}_1 = \mathbf{-0.053}$ $(\boldsymbol{t} = \mathbf{-0.053/\ .141 = 0.38})$.

Now we estimate Model 5,
which is Model 4 modified to have an unstructured covariance matrix.

	Model 4		Model 5	
Fixed effect	Coefficient	S.E.	Coefficient	S.E.
μ_1 Effect of age 55	6.842	0.055	6.842	0.055
μ_2 Effect of age 56	6.914	0.055	6.913	0.056
μ_3 Effect of age 57	6.988	0.055	6.987	0.055
μ_4 Effect of age 58	6.886	0.057	6.882	0.059
μ_5 Effect of age 59	6.892	0.060	6.886	0.060
μ_6 Effect of age 60	7.005	0.064	7.001	0.060
α Main effect birth year	–0.394	0.078	–0.395	0.078
γ Interaction birth year $\times$ age	0.049	0.019	0.043	0.019
Deviance	21,722.45			21,676.84

The covariance matrices implied by Models 4 and 5 are

Model 4 Model 5

$$\begin{pmatrix} 3.67 & 2.16 & 2.06 & 1.96 & 1.87 & 1.77 \\ 2.16 & 3.50 & 2.01 & 1.94 & 1.87 & 1.80 \\ 2.06 & 2.01 & 3.38 & 1.92 & 1.87 & 1.82 \\ 1.96 & 1.94 & 1.92 & 3.31 & 1.87 & 1.85 \\ 1.87 & 1.87 & 1.87 & 1.87 & 3.29 & 1.88 \\ 1.77 & 1.80 & 1.82 & 1.85 & 1.88 & 3.33 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3.61 & 2.11 & 1.92 & 1.92 & 1.84 & 1.60 \\ 2.11 & 3.60 & 2.15 & 1.95 & 1.90 & 1.80 \\ 1.92 & 2.15 & 3.26 & 1.96 & 1.73 & 1.73 \\ 1.92 & 1.95 & 1.96 & 3.49 & 1.85 & 1.85 \\ 1.84 & 1.90 & 1.97 & 2.07 & 3.29 & 1.87 \\ 1.60 & 1.80 & 1.73 & 1.85 & 1.87 & 2.88 \end{pmatrix}$$

Although Model 5 is significantly better than Model 4
($\chi^2 = 45.6,\ df = 15 - 4 = 11, p < 0.001$),
the differences are not important.

15.2 Variable occasions

This is a model of *populations of curves*.

Again, there are random slopes for time and non-linear functions of time.

If there are no variables representing individual $\times$ time interactions,
the fixed effects represent the population averages
and the random effects at level two represent
the between-individual differences in the curves.

Example : height of children with retarded growth, 5–10 years old.

1,886 measurements of a total of 336 children.
Height in cm, age in years, intercept at $t_0 = 5$ years.

The time variable is age; children were measured at quite varying ages.

First a model of linear growth.

Table 15.7 Linear growth model for 5–10-year-old children with retarded growth.

Fixed Effect		Coefficient	S.E.
γ_{00}	Intercept	96.32	0.285
γ_{10}	Age	5.53	0.08
Random Effect		**Parameter**	**S.E.**
Level-two (i.e., individual) random effects:			
τ_0^2	Intercept variance	19.79	1.91
τ_1^2	Slope variance for age	1.65	0.16
τ_{01}	Intercept-slope covariance	–3.26	0.46
Level-one (i.e., occasion) variance:			
σ^2	Residual variance	0.82	0.03
Deviance		7099.87	

Note the small residual standard deviation, $\hat{\sigma} = 0.9$,
which includes measurement error and deviations from linearity.

It follows that growth in this age range is quite nearly linear,
with an average growth rate of 5.53 cm/y,
and a between-individual standard deviation in growth rate of
$\hat{\tau}_0 = \sqrt{1.65} = 1.3$ cm/y.

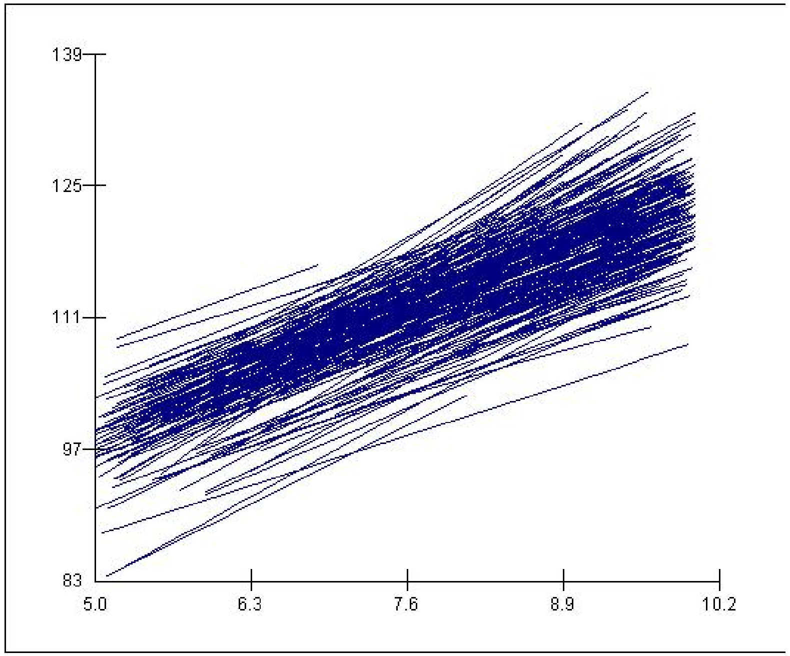
Negative slope-intercept variance:
children who are taller at age 5 grow more rapidly (correlation = –0.63).

Correlation between heights at ages 5 and 10 years is

$$\frac{\text{cov}(U_{0i}, U_{0i} + 5U_{1i})}{\sqrt{\text{var}(U_{0i}) \times \text{var}(U_{0i} + 5U_{1i})}}$$

$$= \frac{19.79 - 5 \times 3.26}{\sqrt{19.79 \times (19.79 - 10 \times 3.26 + 25 \times 1.65)}} = \frac{3.49}{5.33} = 0.65.$$

The ‘predicted’ (postdicted?) curves for the children,
calculated by using empirical Bayes estimates of the parameters:



Next to linear models, polynomial models and spline models can be considered.
Here a simple approach to splines: fixed nodes determined by trial and error.

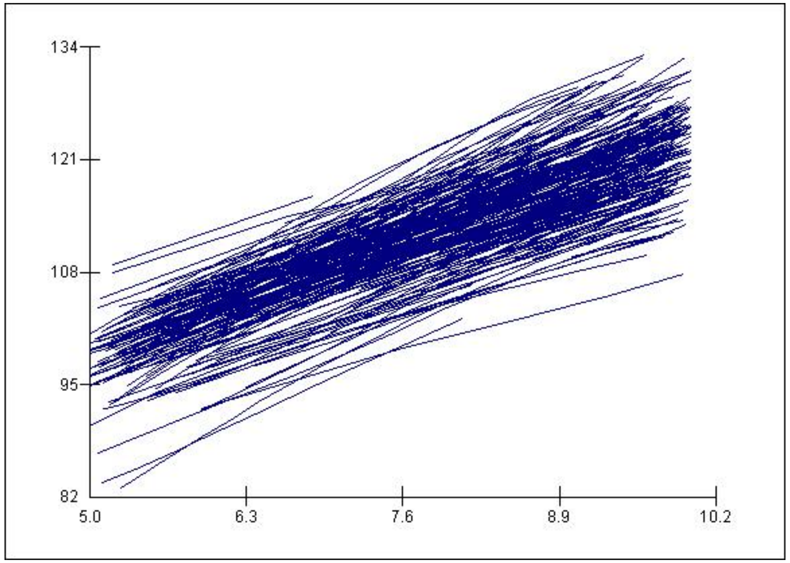
Table 15.8 Cubic growth model for 5–10-year-old children with retarded growth

Fixed Effect		Coefficient	S.E.
γ_{00}	Intercept	110.40	0.22
γ_{10}	$t - 7.5$	5.23	0.12
γ_{20}	$(t - 7.5)^2$	-0.007	0.038
γ_{30}	$(t - 7.5)^3$	0.009	0.020
Random Effect		**Variance**	**S.E.**
Level-two (i.e., individual) random effects:			
τ_0^2	Intercept variance	13.80	1.19
τ_1^2	Slope variance $t - t_0$	2.97	0.32
τ_2^2	Slope variance $(t - t_0)^2$	0.255	0.032
τ_3^2	Slope variance $(t - t_0)^3$	0.066	0.009
Level-one (i.e., occasion) variance:			
σ^2	Residual variance	0.37	0.02
Deviance		6603.75	

Estimated correlation matrix for level-two random slopes $(\boldsymbol{U_{0i}}, \boldsymbol{U_{1i}}, \boldsymbol{U_{2i}}, \boldsymbol{U_{3i}})$:

$$\widehat{\boldsymbol{R}}_U = \begin{pmatrix} \mathbf{1.0} & \mathbf{0.17} & \mathbf{-0.27} & \mathbf{0.04} \\ \mathbf{0.17} & \mathbf{1.0} & \mathbf{0.11} & \mathbf{-0.84} \\ \mathbf{-0.27} & \mathbf{0.11} & \mathbf{1.0} & \mathbf{-0.38} \\ \mathbf{0.04} & \mathbf{-0.84} & \mathbf{-0.38} & \mathbf{1.0} \end{pmatrix} .$$

Again ‘predicted’ curves from empirical Bayes estimates – looks almost linear:



Another option is to fit *piecewise linear function* :
continuous functions of age that can have different slopes
for each interval between birthdays years t and $t + 1$.

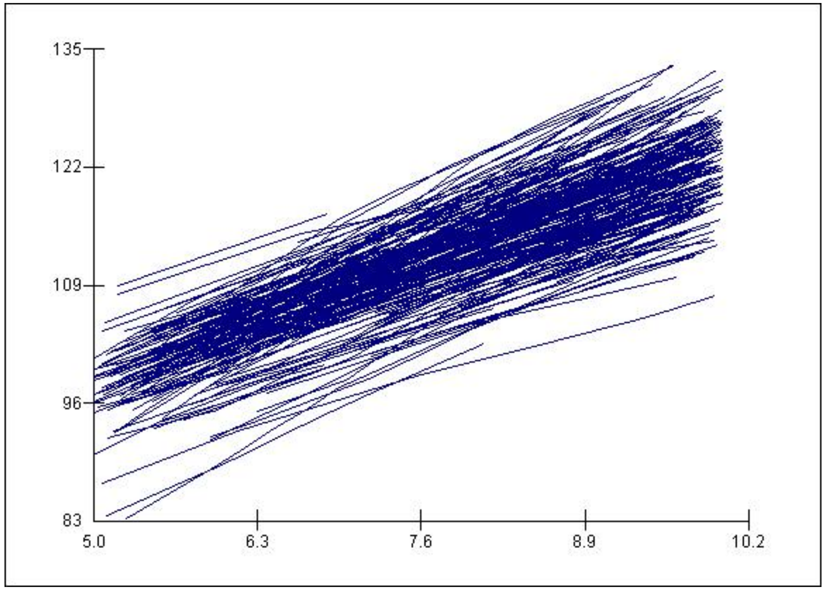
Table 15.9 Piecewise linear growth model for 5–10-year-old children with retarded growth

Fixed Effect		Coefficient	S.E.
γ_{00}	Intercept	110.40	0.22
γ_{10}	f_1 (5–6 years)	5.79	0.24
γ_{20}	f_2 (6–7 years)	5.59	0.18
γ_{30}	f_3 (7–8 years)	5.25	0.16
γ_{40}	f_4 (8–9 years)	5.16	0.15
γ_{50}	f_5 (9–10 years)	5.50	0.16

Random Effect		Variance	S.E.
Level-two (i.e., individual) random effects:			
τ_0^2	Intercept variance	13.91	1.20
τ_1^2	Slope variance f_1	3.97	0.82
τ_2^2	Slope variance f_2	3.80	0.57
τ_3^2	Slope variance f_3	3.64	0.50
τ_4^2	Slope variance f_4	3.42	0.45
τ_5^2	Slope variance f_5	3.77	0.53
Level-one (i.e., occasion) variance:			
σ^2	Residual variance	0.302	0.015
Deviance		6481.87	

Estimated correlation matrix of the level-two random effects $(U_{0i}, \ldots, U_{5i})$ and 'predicted' curves :

$$\widehat{R}_U = \begin{pmatrix} 1.0 & 0.22 & 0.31 & 0.14 & -0.05 & 0.09 \\ 0.22 & 1.0 & 0.23 & 0.01 & 0.18 & 0.33 \\ 0.31 & 0.01 & 1.0 & 0.12 & -0.16 & 0.48 \\ 0.14 & 0.01 & 0.12 & 1.0 & 0.47 & -0.23 \\ -0.05 & 0.18 & -0.16 & 0.47 & 1.0 & 0.03 \\ 0.09 & 0.33 & 0.48 & -0.23 & 0.03 & 1.0 \end{pmatrix} .$$



The advantage of piecewise linear functions is that they are less globally sensitive to local changes in the function values.

The disadvantage is their angular appearance, with kinks in the “knots”.

A further step is to use *splines*,
which are a flexible and smooth family of functions that are like polynomials,
but without the global sensitivity to local changes.

Spline functions are polynomials on intervals bounded by points called *knots*,
and smooth when passing the knots.

We mostly use quadratic or cubic splines.

When the knots have been found with a bit of trial and error,
using the spline function amounts to transforming the time variable
and then further on using *linear* models – see the book.

For example, a cubic spline with one node, positioned at t_0,
is defined by the four functions

$$
\begin{aligned}
f_1(t) &= t - t_0 && \text{(linear function)}\\
f_2(t) &= (t - t_0)^2 && \text{(quadratic function)}\\
f_3(t) &= \begin{cases} (t - t_0)^3 & (t \le t_0) \\ 0 & (t > t_0) \end{cases} && \text{(cubic to the left of } t_0\text{)}\\
f_4(t) &= \begin{cases} 0 & (t \le t_0) \\ (t - t_0)^3 & (t > t_0) \end{cases} && \text{(cubic to the right of } t_0\text{).}
\end{aligned}
$$

Example: growth in the same population of children with retarded growth, now 12-17 year olds. A total of 321 children with 1,941 measurements.

A good model was obtained for cubic splines with one knot at $t_0 = 15$ years.

Table 15.10 Cubic spline growth model for 12–17-year-old children with retarded growth

Fixed Effect		Coefficient	S.E.
γ_{00}	Intercept	150.00	0.42
γ_{10}	f_1 (linear)	6.43	0.19
γ_{20}	f_2 (quadratic)	0.25	0.13
γ_{30}	f_3 (cubic left of 15)	-0.038	0.030
γ_{40}	f_4 (cubic right of 15)	-0.529	0.096
Random Effect		Variance	S.E.
Level-two (i.e., individual) random effects:			
τ_0^2	Intercept variance	52.07	4.46
τ_1^2	Slope variance f_1	6.23	0.71
τ_2^2	Slope variance f_2	2.59	0.34
τ_3^2	Slope variance f_3	0.136	0.020
τ_4^2	Slope variance f_4	0.824	0.159
Level-one (i.e., occasion) variance:			
σ^2	Residual variance	0.288	0.014
Deviance		6999.06	

Estimated correlation matrix of the level-two random effects $(\boldsymbol{U_{0i}}, \dots, \boldsymbol{U_{4i}})$:

$$\widehat{\boldsymbol{R}}_U = \begin{pmatrix} 1.0 & 0.26 & -0.31 & 0.32 & 0.01 \\ 0.26 & 1.0 & 0.45 & -0.08 & -0.82 \\ -0.31 & 0.45 & 1.0 & -0.89 & -0.71 \\ 0.32 & -0.08 & -0.89 & 1.0 & 0.40 \\ 0.01 & -0.82 & -0.71 & 0.40 & 1.0 \end{pmatrix} .$$

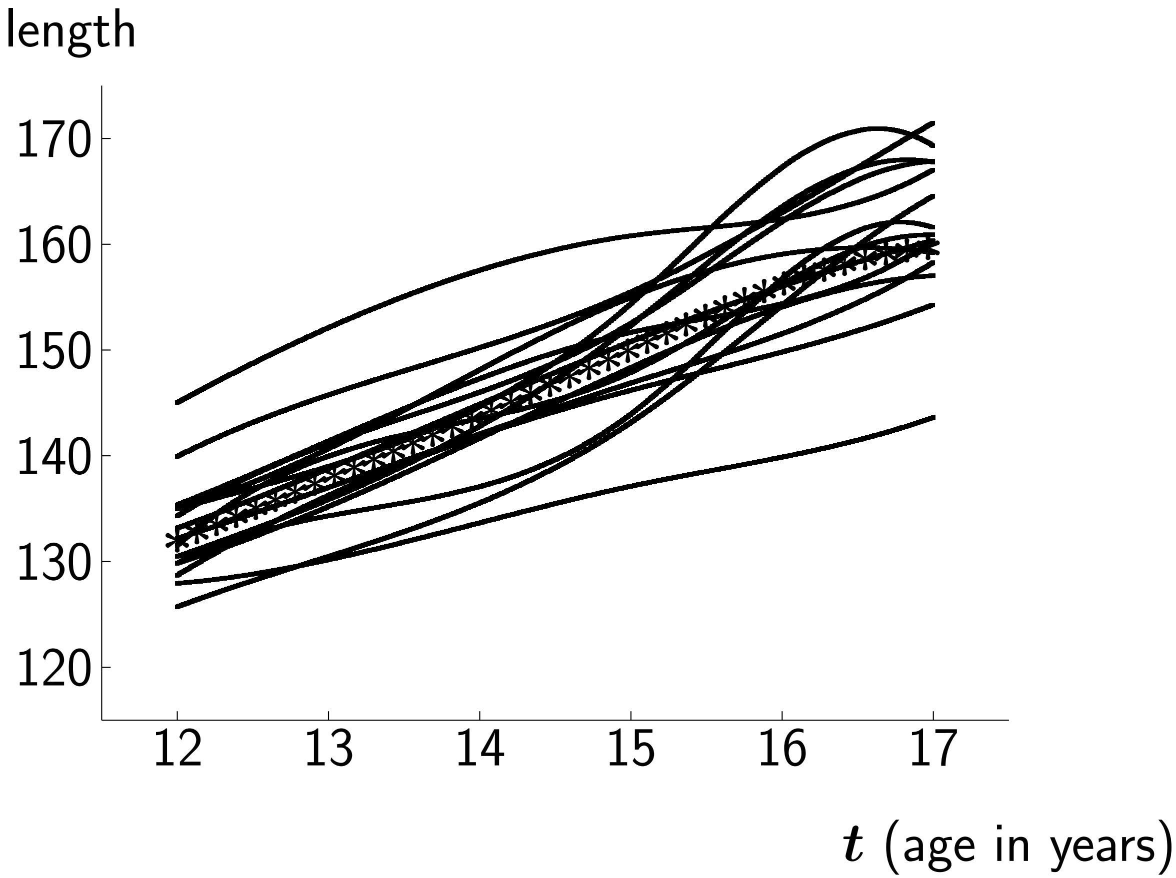


Figure 15.1 Average growth curve ($*$) and 15 random growth curves for 12–17-year-olds for cubic spline model.

Next: explain growth variability by gender and parents' height (Table 15.11).

Fixed Effect		Coefficient	S.E.
γ_{00}	Intercept	150.20	0.47
γ_{10}	f_1 (linear)	5.85	0.18
γ_{20}	f_2 (quadratic)	0.053	0.124
γ_{30}	f_3 (cubic left of 15)	–0.029	0.030
γ_{40}	f_4 (cubic right of 15)	–0.553	0.094
γ_{01}	Gender	–0.385	0.426
γ_{11}	$f_1 \times$ gender	-1.266	0.116
γ_{21}	$f_2 \times$ gender	–0.362	0.037
γ_{02}	Parents' length	0.263	0.071
γ_{12}	$f_1 \times$ parents' length	0.0307	0.0152
Random Effect		Variance	S.E.
Level-two (i.e., individual) random effects:			
τ_0^2	Intercept variance	49.71	4.31
τ_1^2	Slope variance f_1	4.52	0.52
τ_2^2	Slope variance f_2	2.37	0.33
τ_3^2	Slope variance f_3	0.132	0.020
τ_4^2	Slope variance f_4	0.860	0.156
Level-one (i.e., occasion) variance:			
σ^2	Residual variance	0.288	0.013
Deviance		6885.18	

Estimated correlation matrix of the level-two random effects $(\boldsymbol{U_{0i}}, \ldots, \boldsymbol{U_{4i}})$:

$$\widehat{\boldsymbol{R}}_U = \begin{pmatrix} \mathbf{1.0} & \mathbf{0.22} & \mathbf{-0.38} & \mathbf{0.35} & \mathbf{0.07} \\ \mathbf{0.22} & \mathbf{1.0} & \mathbf{0.38} & \mathbf{-0.05} & \mathbf{-0.81} \\ \mathbf{-0.38} & \mathbf{0.38} & \mathbf{1.0} & \mathbf{-0.91} & \mathbf{-0.75} \\ \mathbf{0.35} & \mathbf{-0.05} & \mathbf{-0.91} & \mathbf{1.0} & \mathbf{0.48} \\ \mathbf{0.07} & \mathbf{-0.81} & \mathbf{-0.75} & \mathbf{0.48} & \mathbf{1.0} \end{pmatrix} .$$

The correlation matrix is given only for completeness,
not because you can see a lot from it.

17. Discrete dependent variables

Heterogeneous proportions.

Religious attendance at least once a week, 136,611 individuals in 60 countries.

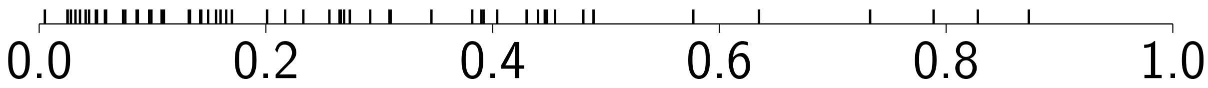


Figure 17.1: Proportion of religious attendance.

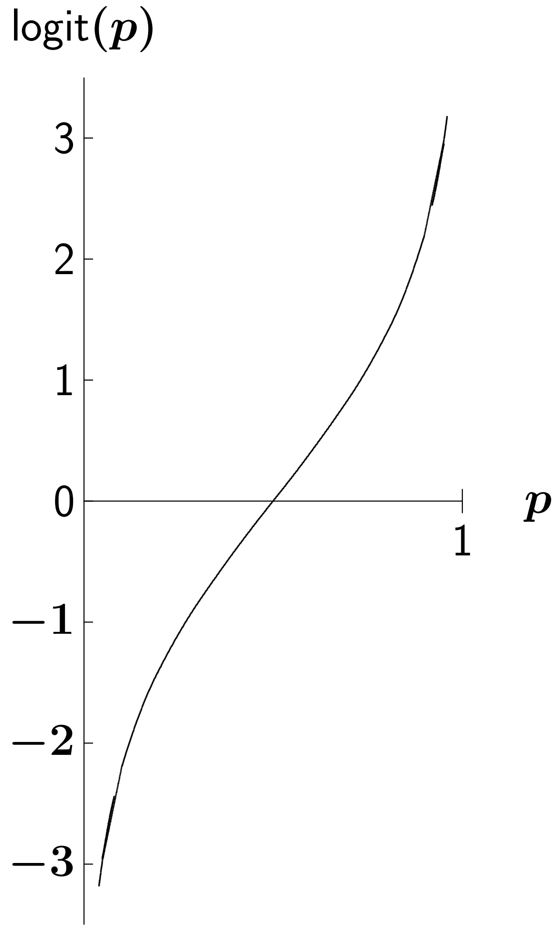
Average proportion 0.238.

Difference between proportions: $\boldsymbol{X^2 = 29{,}733, d.f. = 59, p < 0.0001}$.
$\boldsymbol{\hat{\tau} = \sqrt{0.0404} = 0.201}$.

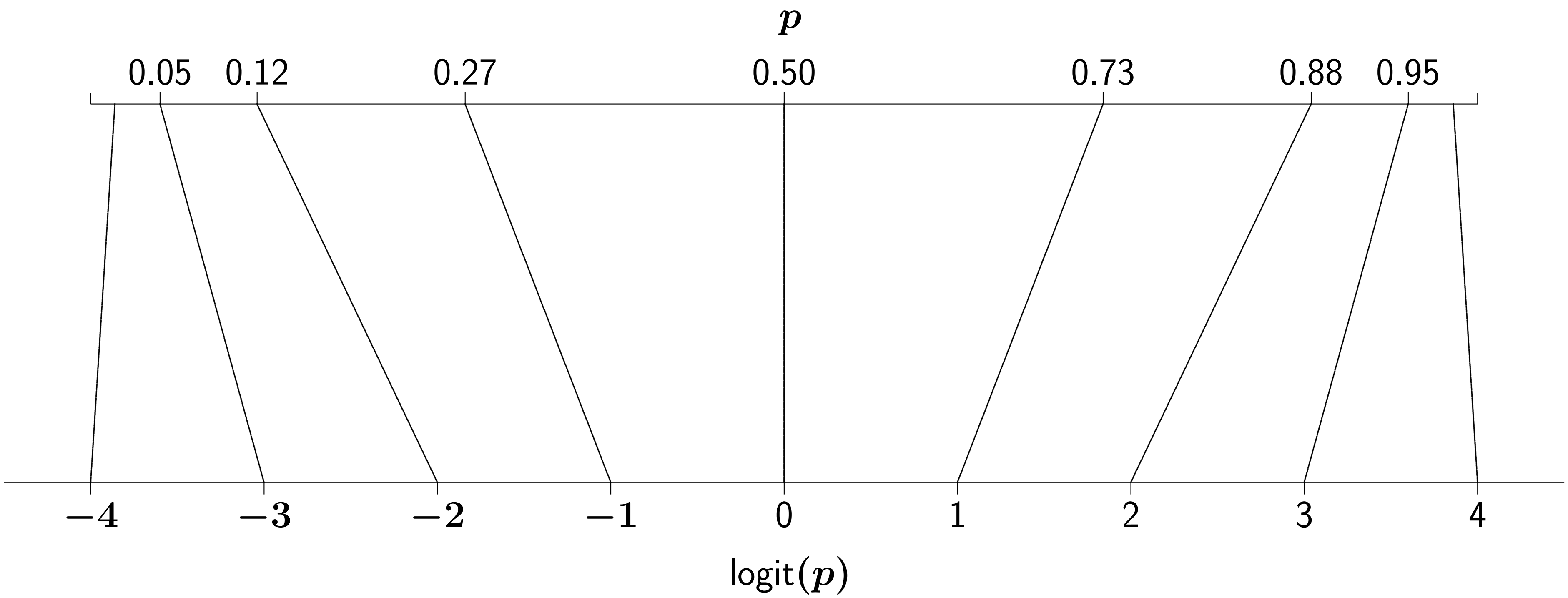
The *logarithm* transforms a multiplicative to an additive scale and transforms the set of positive real numbers to the whole real line. One of the most widely used transformations of probabilities is the *log odds*, defined by

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \;,$$

where $\ln(x)$ denotes the natural logarithm of the number x. The logit function, of which graphs are shown here, is an increasing function defined for numbers between 0 and 1, and its range is from minus infinity to plus infinity.



For example, $p = 0.269$ is transformed to $\text{logit}(p) = -1$ and $p = 0.982$ to $\text{logit}(p) = 4$. The logit of $p = 0.5$ is exactly 0.



Correspondence between p and logit(p).

Empty multilevel logistic regression model: $\text{logit}(P_j) = \gamma_0 + U_{0j}$.

Table 17.1 Estimates for empty multilevel logistic model

Fixed Effect	Coefficient	S.E.
γ_0 = Intercept	-1.447	0.180

Random effect	Var. Comp.	S.E.
Level two variance:		
τ_0 = S.D.(U_{0j})	1.377	

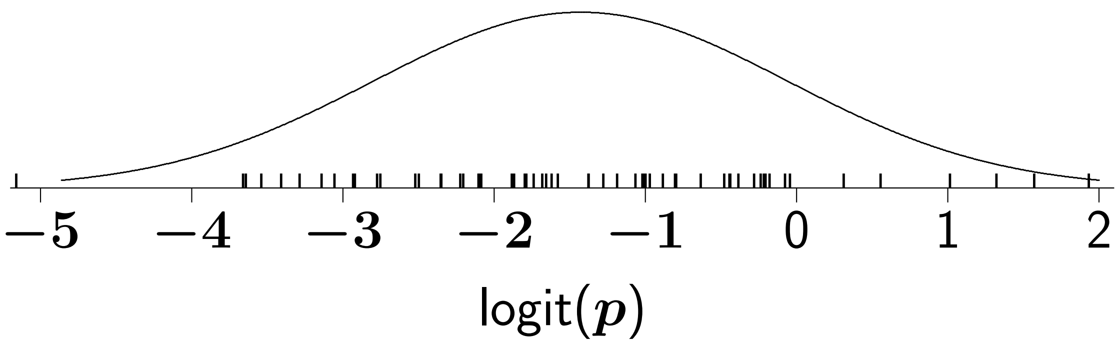


Figure 17.5: Observed log-odds and estimated normal distribution of population log-odds of religious attendance.

The random intercept model for binary data:

$$\text{logit}(P_{ij}) = \gamma_0 + \sum_{h=1}^{r} \gamma_h \, x_{hij} + U_{0j} \; .$$

There is no level-1 residual because P_{ij} is not a random variable, but itself a group-level random coefficient.

Explanatory variables:

- *Educational level*, measured as the age at which people left school (14–21 years), minus 14. This variable was centered within countries. The within-country deviation variable has mean 0, standard deviation 2.48.
- *Income*, standardized within country; mean -0.03, standard deviation 0.99.
- *Employment status*, 1 for unemployed, 0 for employed; mean 0.19, standard deviation 0.39.
- *Sex*, 1 for female, 0 for male; mean 0.52, standard deviation 0.50.
- *Marital status*, 1 for single/divorced/widowed, 0 for married/cohabiting; mean 0.23, standard deviation 0.42.
- *Divorce status*, 1 for divorced, 0 for other; mean 0.06, standard deviation 0.25.
- *Widowed*, 1 for widowed, 0 for other; mean 0.08, standard deviation 0.27.
- *Urbanization*, the logarithm of the number of inhabitants in the community or town of residence, truncated between 1,000 and 1,000,000, minus 10; mean 0.09, standard deviation 2.18.

Constructed level-2 variables:

- Average educational level: mean 0.82, standard deviation 0.94.
- Average unemployment: mean 0.19, standard deviation 0.08.
- Average divorce status: mean 0.06, standard deviation 0.03.
- *Gini coefficient* measuring income inequality, minus 35.
  Mean 0.10, standard deviation 9.58.

Next page:

Table 17.2.
Logistic random intercept model for religious attendance in 59 countries.

Fixed effect	Coefficient	S.E.
γ_0 Intercept	–2.069	0.646
γ_1 Education (within-country deviation)	–0.0290	0.0032
γ_2 Income	–0.0638	0.0082
γ_3 Unemployed	0.017	0.020
γ_4 Female	0.508	0.016
γ_5 Single	–0.269	0.019
γ_6 Divorced	–0.489	0.036
γ_7 Widowed	0.518	0.027
γ_8 Urbanization	–0.0665	0.0039
γ_9 Gini coefficient	0.035	0.017
γ_{10} Country average education	–0.330	0.135
γ_{11} Country average unemployment	6.033	2.116
γ_{12} Country average divorce	–7.120	4.975
Random effect	S.D.	S.E.
Random intercept:		
$\tau_0 = \text{S.D.}(U_{0j})$ intercept standard deviation	1.08	
Deviance	115,969.9	

Random slopes can be added in the usual way
to the multilevel logistic model.

The following is a model for the logit with one random slope:

$$\text{logit}(P_{ij}) = \gamma_0 + \sum_{h=1}^{r} \gamma_h x_{hij} + U_{0j} + U_{1j} x_{1ij} .$$

Table 17.3.
Logistic random slope model for religious attendance in 59 countries.

	Fixed effect	Coefficient	S.E.
γ_0	Intercept	3.792	2.476
γ_1	Education (within-country deviation)	–0.0388	0.0092
γ_2	Income	–0.0738	0.0161
γ_3	Unemployed	0.019	0.020
γ_4	Female	0.511	0.016
γ_5	Single	–0.271	0.019
γ_6	Divorced	–0.493	0.036
γ_7	Widowed	0.482	0.027
γ_8	Urbanization	–0.0650	0.0040
γ_9	Gini coefficient	0.028	0.017
γ_{10}	Country average education	–0.333	0.132
γ_{11}	Country average unemployment	5.44	2.06
γ_{12}	Country average divorce	–6.43	4.84

(continued...)

(... continuation)

Random part parameters		S.D. / Corr.
$\tau_0 = \text{S.D.}(U_{0j})$	Intercept standard deviation	1.09
$\tau_1 = \text{S.D.}(U_{1j})$	Income-slope standard deviation	0.096
$\tau_2 = \text{S.D.}(U_{2j})$	Education-slope standard deviation	0.063
$\rho_{01} = \rho(U_{0j}, U_{1j})$	Intercept–income slope correlation	0.29
$\rho_{02} = \rho(U_{0j}, U_{2j})$	Intercept–education slope correlation	–0.07
$\rho_{12} = \rho(U_{1j}, U_{2j})$	Income–education slopes correlation	0.27

For a data set with large groups like these country data, a two-step approach (Section 3.7) might be preferable.

Estimated level-two intercept variance may go up when level-1 variables are added and always does when these have no between-group variance.

This can be understood by threshold model
which is equivalent to logistic regression:

$$Y = \begin{cases} 0 & \text{if } \breve{Y} \leq 0 \\ 1 & \text{if } \breve{Y} > 0 \,, \end{cases}$$

where $\breve{Y}$ is a latent continuous variable

$$\breve{Y}_{ij} = \gamma_0 + \sum_{h=1}^{r} \gamma_h \, x_{hij} + U_{0j} + R_{ij}$$

and R_{ij} has a logistic distribution, with variance $\pi^2/3$.

The fact that the latent level-1 variance is fixed implies that
explanation of level-1 variation by a new variable X_{r+1}
will be reflected by increase of γ_h $(0 \leq h \leq r)$ and of var(U_{0j}).

Measure of explained variance (‘R^2’) for multilevel logistic regression can be based on this threshold representation, as the
proportion of explained variance in the latent variable.

Because of the arbitrary fixation of σ_R^2 to $\pi^2/3$,
these calculations must be based on one single model fit.

Let

$$\hat{Y}_{ij} = \gamma_0 + \sum_{h=1}^{r} \gamma_h \, x_{hij}$$

be the latent linear predictor; then

$$\breve{Y}_{ij} = \hat{Y}_{ij} + U_{0j} + R_{ij} \, .$$

Calculate $\hat{Y}_{ij}$ (using estimated coefficients) and then

$$\sigma_F^2 = \text{var}\left(\hat{Y}_{ij}\right)$$

in the standard way from the data; then

$$\text{var}\left(\breve{Y}_{ij}\right) = \sigma_F^2 + \tau_0^2 + \sigma_R^2$$

where $\sigma_R^2 = \pi^2/3 = 3.29$.

The proportion of explained variance now is

$$R^2_{\text{dicho}} = \frac{\sigma_F^2}{\sigma_F^2 + \tau_0^2 + \sigma_R^2} \,.$$

Of the unexplained variance, the fraction

$$\frac{\tau_0^2}{\sigma_F^2 + \tau_0^2 + \sigma_R^2}$$

is at level 2, and the fraction

$$\frac{\sigma_R^2}{\sigma_F^2 + \tau_0^2 + \sigma_R^2}$$

is at level one.

Table 17.4 Estimates for probability to take a science subject

	Model 1	
Fixed Effect	Coefficient	S.E.
γ_0 Intercept	2.487	0.110
γ_1 Gender	–1.515	0.102
γ_2 Minority status	–0.727	0.195
Random Effect	Var. Comp.	S.E.
Level-two variance:		
$\tau_0^2 = \text{var}(U_{0j})$	0.481	0.082
Deviance	3238.27	

Linear predictor

$$\hat{Y}_{ij} = 2.487 \; - \; 1.515 \, \text{gender}_{ij} \; - \; 0.727 \, \text{minority}_{ij}$$

has variance $\hat{\sigma}_F^2 = 0.582$. Therefore $R^2_{\text{dicho}} = \frac{0.582}{0.582 + 0.481 + 3.29} = 0.13$.

Multicategory ordinal logistic regression

'Measurement model':

$$\boldsymbol{Y} = \begin{cases} \boldsymbol{0} & \text{if } \breve{\boldsymbol{Y}} \leq \boldsymbol{\theta_0} \\ \boldsymbol{1} & \text{if } \boldsymbol{\theta_0} < \breve{\boldsymbol{Y}} \leq \boldsymbol{\theta_1} \, , \\ \boldsymbol{k} & \text{if } \boldsymbol{\theta_{k-1}} < \breve{\boldsymbol{Y}} \leq \boldsymbol{\theta_k} \;\; (\boldsymbol{k = 2, ..., c - 2}), \\ \boldsymbol{c - 1} & \text{if } \boldsymbol{\theta_{c-2}} < \breve{\boldsymbol{Y}} \, . \end{cases}$$

'Structural model':

$$\breve{\boldsymbol{Y}}_{\boldsymbol{ij}} = \boldsymbol{\gamma_0} + \sum_{\boldsymbol{h=1}}^{\boldsymbol{r}} \boldsymbol{\gamma_h} \, \boldsymbol{x_{hij}} + \boldsymbol{U_{0j}} + \boldsymbol{R_{ij}} \, .$$

Table 17.6 Multilevel 4-category logistic regression model number of science subjects

		Model 1		Model 2	
Threshold parameters		Threshold	S.E.	Threshold	S.E.
$\boldsymbol{\theta_1}$	Threshold 1 - 2	1.541	0.041	1.763	0.045
$\boldsymbol{\theta_2}$	Threshold 2 - 3	2.784	0.046	3.211	0.054
Fixed Effects		Coefficient	S.E.	Coefficient	S.E.
γ_0	Intercept	1.370	0.057	2.591	0.079
γ_1	Gender girls			-1.680	0.066
γ_2	SES			0.117	0.037
γ_3	Minority status			-0.514	0.156
Level two random effect		Parameter	S.E.	Parameter	S.E.
$\boldsymbol{\tau_0^2}$	Intercept variance	0.243	0.034	0.293	0.040
deviance		9308.8		8658.2	

Multilevel Poisson regression

$$\ln\big(E(L_{ij})\big) = \gamma_0 + \sum_{h=1}^{r} \gamma_h \, x_{hij} + U_{0j} \, .$$

L_{ij} is a count variable.

For overdispersed counts, negative binomial models
may be more appropriate than Poisson models;
or a Poisson-type model with overdispersion parameter may be used.

Example next page:
Number of memberships of voluntary organizations of individuals in 40 regions in The Netherlands (ESS data).

Table 17.7. Two Poisson models for number of memberships.

	Model 1		Model 2	
Fixed effect	Coefficient	S.E.	Coefficient	S.E.
γ_0 Intercept	0.846	0.026	0.860	0.032
γ_1 Female			–0.118	0.044
γ_2 $(\text{Age} - 40)/10$			0.198	0.028
γ_3 $(\text{Age} - 40)^2/100$			–0.061	0.014
γ_4 Protestant			0.296	0.041
γ_5 Female $\times$ $(\text{Age} - 40)/10$			–0.097	0.030
Level-two random part	S.D.		S.D. / Corr.	
τ_0 Intercept standard deviation	0.112		0.070	
τ_1 Slope S.D. Female			0.146	
$\rho_{01}(\tau)$ Int.–slope correlation			0.168	
Deviance	6,658.3		6,522.6	

The total contribution of being female is $-0.118 - 0.097(\text{Age} - 40)/10 + U_{1j}$,

which is 0 for age 28 if $U_{1j} = 0$.

Table 17.8. Two negative binomial models for number of memberships.

	Model 3		Model 4	
Fixed effect	Coefficient	S.E.	Coefficient	S.E.
γ_0 Intercept	0.848	0.027	0.861	0.034
γ_1 Female			–0.116	0.046
γ_2 (Age $-$ 40)/10			0.197	0.031
γ_3 (Age $-$ 40)^2/100			–0.061	0.016
γ_4 Protestant			0.300	0.047
γ_5 Female $\times$ (Age $-$ 40)/10			–0.095	0.033
Level-two random part	Parameter		Parameter	
τ_0 Intercept standard deviation	0.111		0.051	
τ_1 Slope S.D. Female			0.127	
$\rho_{01}(\tau)$ Int.–slope correlation			0.497	
α Negative binomial parameter	7.40		10.32	
Deviance	6,588.6		6,483.3	