Notation Survey

Paul E. Johnson\textsuperscript{1} \textsuperscript{2}

\textsuperscript{1}Department of Political Science

\textsuperscript{2}Center for Research Methods and Data Analysis, University of Kansas

2015
Outline

1. Raudenbush and Bryk
2. Snijders and Bosker
3. Rabe-Hesketh & Skrondal
4. Henderson
Outline

1. Raudenbush and Bryk
2. Snijders and Bosker
3. Rabe-Hesketh & Skrondal
4. Henderson
Most people would call this a “Slopes-as-outcomes” (or “Intercepts-as-outcomes” approach)
Basic model

- Individual-level model, dubbed the “Level-1” model.

\[ Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \]

- \( j \): index of the grouping variable
- \( i \): index of the individual observation.
- \( X_{ij} \): The predictor of the participant \( i \) in group \( j \)
- \( \beta_{0j} \): The intercept differs among groups. There is no “constant constant” in this style, it is written down as a group-level coefficient.
- \( \beta_{1j} \): The slope differs across groups as well.
- \( r_{ij} \): The random error term, for the \( i \)'th person in the \( j \)'th group.

- Note in this model, there is no “fixed effect” (no intercept), apart from the group level coefficients.
Level 2 Models

Write out the full story for each j-level thing that varies.

1. A model for the intercept. For each grouping value, $j$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$

- $W_j$ A group-level characteristic.
- $\gamma_{00}$ A constant coefficient, shared among all groups. (“gamma sub zero zero”). Referred to by subscript “00” (I am not entirely sure why).
- $\gamma_{01}$ A coefficient, shared among all groups (“gamma sub zero one”). It maps the group-level variable $W_j$ into the intercepts
- $u_{0j}$ A random error draw that affects only group $j$.
- Notice the mixture of levels involved. We have some system wide constants, some unique-to-group-j variables, and they are blended together into $\beta_{0j}$.
- If you are interested in building a model that includes predictors of $\beta_{0j}$, this is the place. You are “explaining” the group-level blips.
Level 2 Models ...

2 A model for the slope coefficient

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$

- $W_j$ A group-level characteristic, in this case, same as the intercept model. But it might be a different variable.
- $\gamma_{10}$ A constant coefficient, shared among all groups. Referred to by subscript “00” (I am not entirely sure why).
- $\gamma_{11}$ A slope coefficient, shared among all groups, that maps the group-level predictor $W_j$ into the slope coefficient.
- $u_{1j}$ A random error draw that affects only group $j$. 
Assumptions about those random errors

1. All of the error terms have expected value 0.
   \[ E[r_{ij}] = 0 \]
   and
   \[ E \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]
2. The variance of \( r_{ij} \) is the same for all groups \( j \) and all individuals \( i \).
   \[ \text{Var}[r_{ij}] = \sigma^2 \]
   Hence, the base model has completely “homoskedastic” individual and group level errors.
   Recall \([u_{0j}, u_{1j}]^T\) are the “error terms” in the 2 group level models, one for intercept, one for slope.
3. The Variance Matrix of \( u \) is called \( \mathbf{T} \)
   \[ \text{Var} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \mathbf{T} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \]
Assumptions about those random errors ...

1. The variance of $u_{0j}$ is $\tau_{00}$ ("tau sub zero zero"), the variance of $u_{1j}$ is $\tau_{11}$, and the covariance between $u_{0j}$ and $u_{1j}$ is $\tau_{10}$

4. The individual errors $r_{ij}$ and the group errors $u_{kj}$ are uncorrelated:

$$\text{Cov}(u_{0j}, r_{ij}) = \text{Cov}(u_{1j}, r_{ij}) = 0$$
Some folks are mainly interested in the random intercept component, wondering if their subgroups are noticeably different.

To gauge that, some suggest the Intra-Class Correlation coefficient

\[ \rho = \frac{\tau_{00}}{(\tau_{oo} + \sigma^2)} \]
My editorial comments

1. I wondered why subscripts begin at 0. You don’t usually see that anywhere else.
   - Possibly for same reason my fixed effect in regression usually starts at $\beta_0$

2. Letter $r_{ij}$ for individual error is distracting. Group effects $u_{0j}$ and $u_{1j}$ are understandable.

3. Somehow find myself wishing there were a simpler notation, but remember that “t” stands for $\tau$ makes it easier to remember what $T$ is

4. By the same token, the relationship of group 1’s group-level error to the error in group 2 should be made evident here. Different multi-variate notation is needed.
Multi-levels Become One-level Regression!

- On the surface, it appears as though there are two levels.
- The HLM software asks for the data in the 2 separate levels, one with individual level variables and another set with the group level observations.
- However, when the parameters are all estimated, this all becomes a one-level regression.
- Recall, the “Level 1” model

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

- Replace $\beta_{0j}$ and $\beta_{1j}$ with the assumed formulae

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

- So the obvious re-grouping, arrive at what R&B call the “combined form” model.
Multi-levels Become One-level Regression! ...

\[ Y_{ij} = \gamma_{00} + \gamma_{01}W_j + u_{0j} + \gamma_{10}X_{ij} + \gamma_{11}W_jX_{ij} + u_{1j}X_{ij} + r_{ij} \]

I did not think so in 2007, but now it definitely seems peculiar that all of these letters are floating around.
You can see why they think of this as a GLS estimation problem?

- Regroup to separate observable from unobservable predictors:
  \[ Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \gamma_{11} W_j X_{ij} + \{r_{ij} + u_{0j} + u_{1j} X_{ij}\} \]

- Tempting to just run this regression and be finished! (where I throw all the errors into \(e_{ij}\)):
  \[ Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \gamma_{11} W_j X_{ij} + e_{ij} \]

- The variance of the average of \(Y\) across groups is called
  \[ \Delta_j = \text{parameter variance} + \text{error variance} \]

- The details in B&R, p. 40, I prefer not to go into that notation.
- Why is ordinary regression not sufficient?
You can see why they think of this as a GLS estimation problem? ... 

1. Heteroskedasticity. We asserted a strong form of homoskedasticity for $r_{ij}$, but \textit{the combined error is decidedly heteroskedastic}.
   - Cross-row shared intercept errors $u_{0j}$
   - Error proportional to the individual predictor, $u_{1j}X_{ij}$.

2. We don’t get estimates of
   - $\tau_{kj}$
   - $\sigma^2$
Estimates of the random effects

- Discussed in R&B from the GLS point of view.
- If we do ML estimation, we solve simultaneously for the
  - variances of the random effects $\tau_0, \sigma^2$
  - fixed coefficients $\gamma_0, \gamma_1$, etc
Here is how they discuss calculation of the random effects

- Two kinds of estimates of $u_{0j}$.
  - If we fit a separate regression within each group, (completely un-pooled models) we get one estimate.
  - If we pool all the data, we get another estimate.

- The calculation of the group level estimates of $u_{0j}$ and $u_{1j}$ is treated as a post estimation step. ("Empirical Bayes" calculation of estimates of $\beta_{0j}$).

- The discussion is an alphabet soup of coefficient labels flying about. End result is something like

  \[ \text{weight} \cdot \{\text{pooled data estimate}\} + (1 - \text{weight}) \{\text{group specific estimate}\} \]

- The weight depends on the clarity of information within the group.
The HLM software, prepared by R&B, is popular among educational researchers. The book has quite a few estimated model examples. Many are based on the High School and Beyond data set, which we are also using.
One of the models emphasized in R&B is about math achievement and socio economic status. The model supposes that it is not a student’s SES, per se, that is important, but the student’s SES relative to the group in which she is situated.

The group-level average of a predictor is called $\overline{X}_j$. Think of the individual’s difference from that as a measurable predictor.

**Level 1:**

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_j) + r_{ij}$$

And the Level 2 models that go with

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

The essence here is that the effect of the school-centered predictor $X$ has an amount that is common among schools ($\gamma_{10}$) and a part that is unique among schools ($u_{1j}$).
Unobserved Group Level Predictors …

- The combined model

\[ Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_j) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_j) + r_{ij} \]
The Benefits of this Way of Thinking

- I don’t endorse the lettering scheme used here, but
- There is some benefit in writing out your assumptions about the level 2 process and then merging the two into the estimation model.
- The estimation of the regression model always involves interaction effects. (Bad sign: Some aspects of interaction are very poorly understood, even without the clustered random effects.)
Outline

1. Raudenbush and Bryk

2. Snijders and Bosker

3. Rabe-Hesketh & Skrondal

4. Henderson
The Intercept Only Notation

- The individual “empty” (“intercept only”) model (p. 47)

\[ Y_{ij} = \gamma_{00} + U_{0j} + R_{ij} \]

- \(i\) an index for the individual data row
- \(j\) index for the grouping variable
- \(\gamma_{oo}\) The “general mean” of \(Y\), general across groups and individuals
- \(U_{0j}\) A random effect at the level of group \(j\)
- \(R_{ij}\) The random effect different for each data row (for each \(i\) and \(j\))
- \(E[U_{0j}] = 0\)
- \(E[R_{ij}] = 0\)
- \(Var(U_{0j}) = \tau^2_0\) variance of group level effect
- \(Var(R_{ij}) = \sigma^2\) variance of individual effect

- You have a choice in these models. If you include an “intercept”, then the group level effects are seen as variations around that value.
Raudenbush & Bryk would rather say there is no system wide constant, and instead the fundamental thing they are interested in is the group intercept.

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

Interesting point (p. 49).

Consider 2 data rows from same group. Because their individual errors are hypothesized to be uncorrelated, then the variance they share depends solely on $U_{0j}$. Hence the covariance between two rows $i$ and $i'$ within group $j$ is

$$\text{cov}(Y_{ij}, Y_{k'j}) = \text{Var}(U_{0j}) = \tau_0^2$$

The ICC thus reflects the similarity of $i$ and $i'$ within a group.

$$\rho(Y_{ij}, Y_{i'j}) = \frac{\tau_0^2}{\tau_0^2 + \sigma^2}$$
The Intercept Only Notation ...

- ICC, “It can be interpreted in two ways: it is the correlation between two randomly drawn individuals in one randomly drawn group, and it is also the fraction of total variability that is due to the group level”.(p. 50)

- If we introduce a predictor that has the same effect on each individual, none of that analysis changes. Write it out.

\[ Y_{ij} = \gamma_{00} + \gamma_{10}x_{1ij} + \gamma_{20}x_{2ij} \ldots + \gamma_{p0}x_{pij} + U_{0j} + R_{ij} \]

If I were writing this, I’d write \( x_{1ij} \) (the predictor \( x1 \) for row \( i \) within \( j \)). They are half-way to my kind of notation, putting a designated variable name like \( x_1 \) onto which they append subscripts. \( \gamma_1 \) individual level predictors, and the coefficients count up from \( \gamma_{10} \) to \( \gamma_{p0} \).
The Intercept Only Notation ...

- If there are predictors common among members of group $j$, S&B, use the symbol $z_{1j}$, $z_{2j}$, etc. No need for subscript $i$ because these are same among group members $j$. There are $q$ of these group level variables.

$$Y_{ij} = \gamma_{00} + \gamma_{10}x_{1ij} + \gamma_{20}x_{2ij} \cdots + \gamma_{p0}x_{pij}$$
$$+ \gamma_{01}z_{1j} + \gamma_{02}z_{2j} \cdots + \gamma_{0q}z_{3q}$$
$$+ U_{0j} + R_{ij}$$

The coefficients on the variables that are common within group count up, from $\gamma_{01}$ to $\gamma_{0q}$.

- So at least we can tell one thing. The coefficients they want to estimate are called “gamma” with subscripts.

- We’d like to estimate the group-level random effect.

- Are we interested primarily in
The Intercept Only Notation ...

1. $U_{0j}$ or,
2. $\beta_{0j} = \gamma_{00} + U_{0j}$?
   I think the answer is entirely up to you.

- We expect the “true” but unknown $\beta_{0j}$ is somewhere between the
  - estimated group average of $Y_{ij}$, dubbed $\hat{\beta}_{0j} = \overline{Y_{ij}}$ and
  - general mean of $Y_{ij}$, dubbed $\hat{\gamma}_{00}$

- “The optimal combined ’estimate’ for $\beta_{0j}$ is a weighed average of the
  two previous estimates,

$$\hat{\beta}_{0j}^{EB} = \lambda_j \hat{\beta}_{0j} + (1 - \lambda_j) \hat{\gamma}_{00}$$
The superscript “EB” means “empirical Bayes”. The weight looks quite a bit like the ICC

\[ \lambda_j = \frac{\tau_0^2}{\tau_0^2 + \sigma^2/n_j} \]

Different from ICC because as the group size \( n_j \) grows, the role of individual random error shrinks to 0 and \( \lambda_j \to 1 \). That is, as the group grows larger and larger, the weight placed on the group-level information becomes heavier and heavier.
Interesting on the "between group" and "within group" difference

Here is an example of how confusing it becomes when you are looking for a simple rule of thumb.

p. 57. Discussion is about predicting allowance children receive as a function of a child’s age and the average age of children in the family. $x_{ij}$ is the age of the $i$th child in family $j$.

The central question is whether we should estimate the model that uses the child’s age only

$$Y_{ij} = \gamma_0 + \gamma_{10}x_{ij} + U_{0j} + R_{ij}$$

Or if we should include the average age of children in the family, which they label $z_j = \bar{x}_j$,

$$Y_{ij} = \gamma_0 + \gamma_{10}x_{ij} + \gamma_{01}\bar{x}_{ij} + U_{0j} + R_{ij}$$ (1)
Interesting on the "between group" and "within group" difference ...

“This model is more flexible in that the within-group regression coefficient is allowed to differ from the between-group regression coefficient. This can be seen as follows.
If model (1) is considered within a given group, the terms can be reordered as

\[ Y_{ij} = (\gamma_{00} + \gamma_{01} \bar{x}_{ij} + U_{0j}) + \gamma_{10} x_{ij} + R_{ij} \]

The part in parentheses is the (random) intercept for this group, and the regression coefficient of X within the group is \( \gamma_{10} \). The systematic (nonrandom) part for group j is the within-group regression line

\[ Y = (\gamma_{00} + \gamma_{01} \bar{x}_{j}) + \gamma_{10} x. \]
Interesting on the "between group" and "within group" difference ...

On the other hand, taking the group average on both sides of the equality sign in (1) yields the between-group regression model,

$$\bar{Y}_j = \gamma_{00} + \gamma_{10} \bar{x}_j + \gamma_{01} \bar{x}_j + U_{0j} + \bar{R}_j$$

$$= \gamma_{00} + (\gamma_{10} + \gamma_{01}) \bar{x}_j + U_{0j} + \bar{R}_j$$

The systematic part of this model is represented by the between-group regression line

$$Y = \gamma_{00} + (\gamma_{10} + \gamma_{01}) x.$$ 

This shows that the between-group regression coefficient is \(\gamma_{10} + \gamma_{01}\). The difference between the within group and between-group regression coefficients can be tested in this model by testing the null hypothesis that \(\gamma_{01} = 0\)...
Interesting on the "between group" and "within group" difference ...

I Wish the subscripts go missing on \( Y \) in two of those, so I'm not exactly sure what it is saying. They continue on the next page (p. 58).

"If the within and between-group regression coefficients are different, then it is often convenient to replace \( x_{ij} \) in (1) with the within-group deviation score, defined as \( x_{ij} - \bar{x}_j \). To distinguish the corresponding parameters from those in (1), they are denoted by \( \tilde{\gamma} \). The resulting model is

\[
Y_{ij} = \tilde{\gamma} + \tilde{\gamma}(x_{ij} - \bar{x}_j) + \tilde{\gamma}_{01}\bar{x}_j + U_{0j} + R_{ij}
\]

This model is statistically equivalent to model (1), but has a more convenient parameterization because the between-group regression coefficient is now

\[
\tilde{\gamma}_{01} = \gamma_{10} + \gamma_{01}
\]
Interesting on the "between group" and "within group" difference ...

while the within-group regression coefficient is

\[ \tilde{\gamma}_{10} = \gamma_{10}. \]

The conclusion I draw from this is that I don’t care which model is fit, they are the same. It is the same as the general “mean-centering” tragedy that is an on-going hassle.

Now turn to p. 88, where we find out that those models are NOT equivalent if there is a random slope involved.

(p. 88)

“In Section 4.6, a model was introduced by which we could distinguish within- from between-group regression. These models were discussed:

\[ Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}\bar{x}_{ij} + U_{0j} + R_{ij} \] (2)
Interesting on the "between group" and "within group" difference ...

and

\[ Y_{ij} = \gamma + \gamma (x_{ij} - \bar{x}_j) + \gamma_0 \bar{x}_j + U_{0j} + R_{ij} \]  \hspace{1cm} (3)

It was shown that \( \gamma_0 = \gamma_{10} + \gamma_{01}, \gamma_{10} = \gamma_{10} \), so that the two models are equivalent.

Are the models also equivalent when the effect of \( X_{ij} \) or \( (X_{ij} - \bar{X}_j) \) is random across groups? This was discussed by Kreft et al. (1995). Let us first consider the extension of (2). Define the level-one and level-two models

\[ Y_{ij} = \beta_0 + \beta_{1j} x_{ij} + \gamma_0 \bar{x}_j + R_{ij} \]
\[ \beta_{0j} = \gamma_{00} + U_{0j} \]
\[ \beta_{1j} = \gamma_{10} + U_{1j} \]
Interesting on the "between group" and "within group" difference ...

substituting the level-two model into the level-one model leads to

\[ Y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \gamma_{01}\bar{x}_j + U_{0j} + U_{1j}x_{ij} + R_{ij} \]

Next we consider the extension of (3):

\[ Y_{ij} = \tilde{\beta}_{0j} + \tilde{\beta}_{1j}(x_{ij} - \bar{x}_j) + U_{0j} + U_{1j}x_{ij} + R_{ij} \]

\[ \tilde{\beta}_{0j} = \tilde{\gamma}_{00} + U_{0j} \]

\[ \tilde{\beta}_{1j} = \tilde{\gamma}_{10} + U_{1j} \]

substitution and replacement of terms now yields

\[ Y_{ij} = \tilde{\gamma}_{00} + \tilde{\gamma}_{10}x_{ij} + (\tilde{\gamma}_{01} - \tilde{\gamma}_{10})\bar{x}_j + U_{0j} + U_{1j}x_{ij} - U_{1j}\bar{x}_j + R_{ij} \]
Interesting on the "between group" and "within group" difference ...

This shows that the two models differ in the term $U_1j\bar{x}.j$ which is included in the group-mean-centered random slope model but not in the other model. Therefore in general there is no one-to-one relation between the $\gamma$ and the $\tilde{\gamma}$ parameters, so the models are not statistically equivalent except for the extraordinary case where variable $X$ has no between-group variability.

This implies that in constant slope models one can use either $X_{ij}$ and $\bar{X}.j$ or $(X_{ij} - \bar{X}.j)$ and $\bar{X}.j$ as predictors, since this results in statistically equivalent models, but in random slope models one should carefully choose one or the other specification depending on substantive considerations and/or model fit.

On which consideration should this choice be based? Generally, one should be reluctant to use group-mean centered random slopes unless there is a clear theory (or an empirical clue) that not the absolute score $X_{ij}$ but rather the relative score $(X_{ij} - \bar{X}.j)$ is related to $Y_{ij}$."
Outline

1. Raudenbush and Bryk
2. Snijders and Bosker
3. Rabe-Hesketh & Skrondal
4. Henderson
This is a very well written 2 volume set that has plenty of statistical theory and lots of examples of usage for Stata
Variance components model with only intercept predictors

\[ y_{ij} = \beta + \zeta_j + \varepsilon_{ij} \]

\( i \) is for individual
\( j \) is for group
The “error components” part of the model is

\[ \xi_{ij} = \zeta_j + \varepsilon_{ij} \]

\[ E[\zeta_j] = 0, \ E[\varepsilon_{ij}] = 0 \]

\[ \text{Var}(\zeta_j) = \psi \text{ and } \text{Var}(\varepsilon_{ij}|\zeta_j) = \theta \]
formulation of the model ... 

The variation of the observed $y_{ij}$ among the rows of data is

$$\text{Var}(y_{ij}) = E[(y_{ij} - \beta)^2] = E[(\zeta_j + \epsilon_{ij})^2]$$

$$= E[\zeta_j^2] + 2E[\zeta_j\epsilon_{ij}] + E[\epsilon^2]$$

$$= \psi + 0 + \theta$$

Proportion of total variance that is “between subjects”

$$\rho = \frac{\text{Var}(\zeta_j)}{\text{Var}(y_{ij})} = \frac{\psi}{\psi + \theta}$$

That can be interpreted as the correlation of scores for 2 different observations within a single group, known as the Intraclass Correlation Coefficient.

p. 128.
formulation of the model …

“…we obtain a linear random-intercept model with covariates:

\[ y_{ij} = \beta_1 + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + (\zeta_j + \epsilon_{ij}) \]  
\[ = (\beta_1 + \zeta_j) + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + \epsilon_{ij} \]  

p. 143

\{The between-estimator.\}

“If we wanted to obtain purely between-mother effects of the covariates, we could average the response and covariates for each mother \(j\) over children \(i\) and perform the regression on the resulting means. …

\[ \bar{y}_{ij} = \beta_1 + \beta_2 \bar{x}_{2.j} + \ldots + \beta_p \bar{x}_{p.j} + \zeta_j + \bar{\epsilon}.j \]

…Any information on the regression coefficients from within-mother variability is eliminated, and the coefficients of covariates that do not vary between mothers are absorbed by the intercept.”
formulation of the model ...

p. 152.
“We can easily relax the assumption that the between and within effects are the same for a particular covariate, say, $x_{2ij}$, by using the model:

$$y_{ij} = \beta_1 + \beta_2^W (x_{2ij} - \bar{x}_{2,j}) + \beta_2^B \bar{x}_{2,j} + \beta_3 x_{3ij} + \beta_p x_{p ij} + \zeta_j + \epsilon_{ij}$$

which collapses to the original random-intercept model in (4) if $\beta_2^W = \beta_2^B = \beta_2$.”

p. 164
A linear model can be written as

$$y = X\beta + \xi$$

“In practice, we know only the structure of the residual covariance matrix, assuming that our model is correct. Namely, according to the
formulation of the model ...

random-intercept model, the variance is constant, residuals are correlated within clusters with constant correlation and uncorrelated across clusters.” (p. 164)

The variance-covariance matrix of “all total residuals in the dataset \((\xi_{11}, \xi_{21}, \ldots, \xi_{n_11}, \xi_{12}, \ldots, \xi_{n_J1})\) has block diagonal form:

\[
V = \begin{pmatrix}
V_1 & 0 & 0 & 0 \\
0 & V_2 & 0 & \\
0 & 0 & \ddots & \\
0 & 0 & 0 & V_J
\end{pmatrix}
\]

The blocks \(V_2\) to \(V_J\) on the diagonal are the within-cluster covariance matrices, and all other elements are zero. For a cluster with \(n_j = 3\) units, the within-cluster covariance matrix has the structure

\[
V_j = \begin{bmatrix}
\psi + \theta & 0 & 0 \\
0 & \psi + \theta & 0 \\
0 & 0 & \psi + \theta
\end{bmatrix}
\]
formulation of the model ...” (p. 165).

Estimation methods: GLS based

“In FGLS, the regression coefficients are first estimated by OLS, yielding estimated residuals from which the residual covariance matrix is estimated as $\hat{V}$. The regression coefficients are then reestimated by substituting $\hat{V}$ for $V$ in the GLS estimator, producing estimates $\hat{\beta}_{FGLS}$.... This suggests reestimating $V$ based on the FGLS residuals and then reestimating the regression coefficients. Iterating this process until convergence yields so-called iterative generalized least squares (IGLS).”

“More commonly, the likelihood is maximized using gradient methods such as the Newton-Raphson algorithm”(p. 165)

Random Coefficient model (random slope)

\[
y_{ij} = \beta_1 + \beta_2 x_{ij} + \zeta_{1j} + \zeta_{2j} x_{ij} + \epsilon_{ij} \\
(\beta_1 + \zeta_{1j}) + (\beta_2 + \zeta_{2j}) x_{ij} + \epsilon_{ij}
\]
(p. 190) “Given $X_j$, the random intercept and the random slope have a bivariate distribution assumed to have zero means and covariance matrix $\bar{\psi}$:

$$
\begin{bmatrix}
\psi_{11} & \psi_{12} \\
\psi_{21} & \psi_{22}
\end{bmatrix}
\begin{bmatrix}
\text{Var}(\zeta_{1j}|X_j) & \text{Cov}(\zeta_{1j},\zeta_{2j}|X_j) \\
\text{Cov}(\zeta_{2j},\zeta_{1j}|X_j) & \text{Var}(\zeta_{2j}|X_j)
\end{bmatrix}
$$
Outline

1. Raudenbush and Bryk
2. Snijders and Bosker
3. Rabe-Hesketh & Skrondal
4. Henderson

This appears to be a comprehensive restatement of Henderson’s research, which began as a graduate student at Iowa State in the 1950s. While this was not well known among social scientists until very recently, Henderson pioneered the fundamental ideas of random effects modeling and multi-level analysis in his graduate research seminars. After that, he had a long and thoughtful career as a researcher.
Basic model

\[ y = X\beta + Zu + e \]

- **y**: A vector of \( n \) observations on the outcome
- **X**: fixed observations, \( n \times p \), \( \text{rank}=r \leq \min(n, p) \)
- **\( \beta \)**: \( p \times 1 \) vector to be estimated
- **Z**: a known, fixed, \( n \times q \) matrix
- **u**: a random vector \( q \times 1 \) with “null means” (expected value 0)
- **e**: random vector \( n \times 1 \) with “null means”
Variance Structures

G the variance-covariance matrix of $u$, a $q \times q$ symmetric matrix. Notation: $Var(u)$ is the variance matrix of $u$

R the variance-covariance matrix of $e$, an $n \times n$ symmetric matrix. Assume: $Cov(u, e^T) = 0$, “that is, all elements of the covariance matrix for $u$ with $e$ are zero in most but not all applications.” (p. 2) “Generally we do not know the values of the individual elements of $G$ and $R$. We usually are willing, however, to make assumptions about the pattern of these values. For example, it is often assumed that all the diagonal elements of $R$ are equal and that all off-diagonal elements are zero. That is, the elements of $e$ have equal variances and are mutually uncorrelated.”

The variance of observed $y$ is

$$Var(y) = ZGZ^T + R$$

PJ explanation: Set the GLS objective function as

$$(y - \hat{y})^T Var(y)^{-1} (y - \hat{y})$$

If we could invert $Var(y)$, we could calculate $\hat{\beta}$ by solving the first order conditions. We cannot invert...
Estimation

(Ch 3, p. 2)"One frequent difficult with GLS equations, particular in the mixed model, is that \( ZGZ^T + R \) is large and non-diagonal. Consequently, \( V^{-1} \) is difficult or impossible to compute by usual methods. It was proved by Henderson et al. (1959) that

\[
V^{-1} = R^{-1} - R^{-1}Z(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}
\]

Now if \( R^{-1} \) is easier to compute than \( V^{-1} \), as is often true, if \( G^{-1} \) is easy to compute, and \((Z^TR^{-1}Z + G^{-1})^{-1}\) is easy to compute, this way of computing \( V^{-1} \) may have important advantages. Note that this result can be obtained by writing equations, known as Henderson’s mixed model equations (1950) as follows,

\[
\begin{pmatrix}
X^TR^{-1}X & X^TR^{-1}Z \\
Z^TR^{-1}X & Z^TR^{-1}Z + G^{-1}
\end{pmatrix}
\begin{bmatrix}
\hat{\beta} \\
\hat{u}
\end{bmatrix}
= 
\begin{bmatrix}
X^TR^{-1}y \\
Z^TR^{-1}y
\end{bmatrix}.
\]

This is the famous MME, the derivation of which is difficult.
My emphasis in writing this out is the following. The things we want to know, $\beta$ and $u$, can be thought of as one long vector of coefficients. They can be estimated a part of one piece, in one giant equation that looks quite a bit like the normal equations for a regression. The suggested approach is iterative. Calculate estimates of $\hat{u}$. Then calculate estimates of $\hat{\beta}$. The derivation of this somewhat obscure. It becomes much more understandable in the modern context where it would be derived as a solution to a problem in penalized least squares (Bates and Debroy, Journal of Multivariate Analysis). Henderson originally presented this process at a convention in the 1950s and he called the calculation of $\hat{u}$ “estimation.” Statisticians at the time thought it was better referred to as prediction. Soon after that, Goldberger used the term “Best Linear Unbiased Prediction” (BLUP) which Henderson adopted in his 1984 manuscript. The BLP is the mean of the group-specific random effect,