INTRODUCTION TO MULTILEVEL MODELING

BACKGROUND

A common statistical assumption is that the observations or cases are sampled independently from one another (e.g., via simple random sampling). In practice, however, many samples are generated in stages in which a certain number of primary units are selected and, from these, secondary units are then sampled. The cases are, therefore, not independent.

Even if simple random sampling is used, the cases may not be independent because they may share a 'natural' grouping that is unrelated to the sampling procedure. Another way to say this, using people as an example, is that people are not independent because they are nested within many different clusters (e.g., they live in the same country, metropolitan area, or neighborhood; they work for the same firm, etc.).

There are many other examples of nested data:

- Meta analysis – research studies nested in research methods (e.g., quantitative analyses of prejudice nested in different research methods, such as different measurement strategies)
- Modeling growth – observations nested in individuals (e.g., repeated vocabulary tests nested in students)

Traditional (and incorrect) methods of dealing with hierarchical/nested/multilevel data

Some people are not interested in exploring the effects of the "larger context." For example, someone may be interested in examining the sources of prejudice and they may not care that the ethnic composition of the metropolitan area may affect prejudice. Even though this person makes no attempt to incorporate group-level variables (e.g., the racial composition of the area), their cases (individuals) are not independent if they are clustered by metropolitan areas – and there is a good chance that they will suffer the consequences in their analyses (i.e., correlated errors and heteroskedasticity).

On the other hand, someone may have data for multiple units of analysis – e.g., individual-level data and metropolitan area-level data. For these people, there have been two basic strategies to deal with hierarchical or nested data: disaggregation and aggregation.

Disaggregation – Disaggregation ("pooling the data") means assigning level-2 variables (the higher level) to the level-1 cases. For example, in these data, all cases in the same group have the same score on all group-level variables:

Case	Group	Prejudice (z score)	Education in years	Percent foreign born
1	1	1.34	11	13.1
2	1	1.10	10	13.1
3	2	-1.92	16	8.5
4	2	0.03	12	8.5

There are a number of problems with disaggregating all variables to the lower level:

1. You cannot assume that all cases are independent. Non-independence of cases leads to correlated errors and heteroskedasticity (unequal error variances). The consequences are that the OLS regression slopes are not the minimum variance estimates and the standard errors and, therefore, the t tests are mis-estimated.

- Correlated errors – it is practically impossible to control in a regression equation for all of the similarities between cases in the same group. These similarities that are not controlled disappear into the error term. So, typically, the errors for two cases within the same group will be similar and the errors for two cases in different groups will be dissimilar…thus, there is a systematic relationship between the errors.

- Heteroskedasticity – we may be better able to predict the outcome in some groups than others (the variance in the errors will be smaller for groups for which we are able to predict the outcome well)

2. Also, you should probably not assume that the regression slope for group 1 is equal to the slope for group 2, etc. (e.g., perhaps the relationship between prejudice and education varies by metropolitan area). "Heterogeneity" in regression slopes (differences in the slopes across level-2 units) is common. If you ignore the grouping of level-1 units, you force the relationship (the regression slope) to be the same across all groups. You also force the intercepts (mean levels of the dependent variable) to be the same.

3. Aggregation biases can lead to incorrect conclusions. Group-level variables can be reducible or non-reducible – for example, school SES and school type (Catholic, public). Reducible variables mean very different things when measured at different levels of analysis. SES – at the student level – is an indicator of the resources available at home. School SES (the average student SES in the school) is a measure of the school's resources. Monte Carlo simulations have demonstrated that the effects of group-level, reducible variables are often underestimated when data are disaggregated (Bidwell and Kasarda 1980).

Aggregation – An alternative to assigning all variables to the lower level unit of analysis is to aggregate all variables – in other words, to assign all variables to the higher-level unit of analysis and to use OLS regression. For example, you could examine the effect of percent foreign born on group mean prejudice.

1. By aggregating the data, you throw away a lot of information – all within-group variation is gone. For most of the examples, the within-group variance will comprise 70-90% of the total variance in the outcome. In other words, there is usually more variation across cases in the level of the outcome within groups than variation across groups.

2. The relationships between aggregated variables are usually inflated / overestimated.

3. The relationship between two aggregated variables is often much different than the relationship between "equivalent" variables measured at other units of analysis. For example, in individual-level studies of ethnic and racial prejudice, scholars have demonstrated that inter-ethnic contact reduces prejudice (both variables are measured at the individual level). However, in aggregate studies prejudice is higher in regions with greater opportunities for contact (both variables are measured at the group level).

Multilevel modeling

Multilevel modeling techniques control for the non-independence of cases by including a more sophisticated error term in the regression equation. It also allows you to easily model differences in slopes and intercepts.

Models

There are a variety of different sub-models. These allow you to:

1. Test to see if the mean outcome differs across level-2 units – e.g., does the level of prejudice vary across countries?

2. Estimate regression models with only level-1 independent variables while controlling for the statistical problems often associated with nested data – e.g., regressing prejudice on years of education.

3. Test to see if the effects of level-1 variables differ across level-2 units – e.g., does the effect of education on prejudice vary across countries?

4. Model or explain differences in the average level of the dependent variable across level-2 units – e.g., use country-level variables to explain why the average level of prejudice is higher in some countries.

5. Model or explain differences in the effects of level-1 variables – e.g., use country-level variables to explain why the effect of education on prejudice varies across countries.

THE GENERAL MODEL
1. Education and prejudice in one country:
$Y_i = \beta_0 + \beta_1 X_i + r_i$

- Y_i is the observed value of the dependent variable, prejudice, for respondent i
- β_0 is the intercept – or the predicted value of prejudice when education equals zero
- β_1 is the regression slope for X – or the effect of education on prejudice
- X_i is the observed value of the independent variable, education, for respondent i
- r_i is the prediction error for respondent i – or the difference between the observed and predicted prejudice score

2. Education and prejudice in two countries:
(1) $Y_i = \beta_0 + \beta_1 X_i + r_i$
(2) $Y_i = \beta_0 + \beta_1 X_i + r_i$

The more level-2 units that you have, the more difficult and cumbersome it becomes to estimate separate models for each. Instead of this, we could pool the data and estimate one equation.

Centering: A Quick and Necessary Detour
Centering is useful for simplifying the interpretation of the intercept. Also, a specific centering option is required for some models…we will deal with that later.

The value of the intercept is not always meaningful because it may be impossible to have a score of zero on some independent variables (e.g., age). We can make the intercept more meaningful by centering the independent variable – you can do this by subtracting the mean value of the variable from each person's score:

Age (mean=40)	Mean centered age
38	-2
39	-1
40	0
41	1
42	2

The intercept is still the predicted prejudice when age equals zero. However, the zero value for age is now possible – it is even meaningful because zero is the mean value of age. So the intercept is now the predicted prejudice for a respondent of average age.

NOTE – standardizing a variable will accomplish the same thing, but will change the interpretation because it changes the metric of the variable (e.g., from years of age to standard deviations of age). Centering does not change the interpretation.

You have to create your own centered variables in STATA. Three automated options are available in HLM: no centering, group-mean centering, and grand-mean centering. Group-mean centering: subtract the country mean age from the observed age for all respondents within each country. Grand-mean centering: subtract the mean age (mean across all respondents in all countries) from the observed age for all respondents. For example:

Respondent	Country	Uncentered	Group mean centered	Grand mean centered
1	1	39	-1	-6
2	1	40	0	-5
3	1	41	1	-4
4	2	44	-1	-1
5	2	45	0	0
6	2	46	1	1
7	3	49	-1	4
8	3	50	0	5
9	3	51	1	6

Grand mean=45

3. Education and prejudice in J countries:

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

Two things have changed from the previous equations:

- The addition of the j subscript – i refers to the respondent (where i = 1 to n_j) and j refers to the country (where j = 1 to J).
- X_i was replaced by $(X_{ij} - \overline{X_{.j}})$ – this just indicates that the independent variable, education, is group mean centered.

So…

- Y_{ij} is the observed value of the dependent variable, prejudice, for respondent i in country j
- β_{0j} is the intercept for country j (each country gets its own intercept)
- β_{1j} is the regression slope for X, or education, for country j (each country gets its own slope)
- $(X_{ij} - \overline{X_{.j}})$ is the group-mean centered education score for respondent i in country j
- r_{ij} is the prediction error for respondent i in country j – or the difference between the observed and predicted prejudice for respondent i in country j

The average intercept (the average across all countries) is called γ_0

The average regression slope (the average across all countries) is called γ_1

Both of these things (slopes and intercepts) have variance – that is, they vary across countries. It is assumed that the slopes and intercepts come from a bivariate normal distribution across the population of countries. What's next? Well, if there is variance in the slopes and/or intercepts across countries, then we should try to explain it! For example, we can use characteristics of the countries to explain why the regression slope is more negative/positive in some countries than others or why some countries have higher average levels of the dependent variable.

So we can write regression equations for the intercepts and slopes:

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

- β_{oj} is the intercept for country j
- γ_{00} is the grand mean prejudice (the average intercept across all countries)
- γ_{01} is the effect of W_j (e.g., percent foreign born) on the intercept
- W_j is a country-level independent variable (e.g., percent foreign born)
- u_{oj} is the error or the difference between the observed and predicted intercept for country j

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

- β_{1j} is the education slope for country j
- γ_{10} is the grand mean education slope (the average slope across all countries)
- γ_{11} is the effect of W_j (e.g., percent foreign born) on the education slope
- W_j is a country-level independent variable (e.g., percent foreign born)
- u_{1j} is the error or the difference between the observed and predicted slope for country j

And now…lets put all of the different equations together:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \overline{X_{\cdot j}}) + \gamma_{11} W_j (X_{ij} - \overline{X_{\cdot j}}) + u_{0j} + u_{1j}(X_{ij} - \overline{X_{\cdot j}}) + r_{ij}$$

- Y_{ij} is the observed prejudice for respondent i in country j
- γ_{00} is the grand mean prejudice (the average intercept across all countries controlling for education and percent foreign born)
- γ_{01} is the effect of W_j, percent foreign born, on the intercept
- W_j is a country-level independent variable (e.g., percent foreign born)
- γ_{10} is the grand mean education slope (the average slope across all countries)
- $(X_{ij} - \overline{X_{\cdot j}})$ this is the group mean centered independent variable, education
- γ_{11} is the effect of W_j, percent foreign born, on the education slope

Notice the complicated error structure:

$$u_{0j} + u_{1j}(X_{ij} - \overline{X_{\cdot j}}) + r_{ij}$$

- r_{ij} is the difference between the observed and predicted prejudice for respondent i in country j
- u_{0j} is the difference between the observed and predicted intercept for country j
- u_{1j} is the difference between the observed and predicted education slope for country j

The error is dependent within each country because u_{0j} and u_{1j} are the same for all respondents in country j

The error is unequal across countries (there is heteroskedasticity) because u_{0j} and u_{1j} vary across countries and $(X_{ij} - \overline{X_{\cdot j}})$ varies across respondents.

This is the "general model." We can answer all five questions on pages 2 and 3 by setting different parts of this equation equal to zero – in other words, if we cancel them out.

ESTIMATION BASICS

The General Model

Level 1 Model: $Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X_{.j}}) + r_{ij}$

Level 2 Models: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$

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

Combined Model: $Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \overline{X_{\cdot j}}) + \gamma_{11} W_j (X_{ij} - \overline{X_{\cdot j}}) + u_{0j} + u_{1j}(X_{ij} - \overline{X_{\cdot j}}) + r_{ij}$

Observed variables: Y_{ij}, W_j, X_{ij}, Estimated parameters:

Fixed effects: γ_{00}, γ_{01}, γ_{10}, γ_{11}

Random effects: β_{0j}, β_{1j}

Variance/covariance components:

$var(r_{ij})=\sigma^2$, $var(u_{0j})=\tau_{00}$, $var(u_{1j})=\tau_{11}$, $cov(u_{0j}, u_{1j})=\tau_{01}$

Fixed effects

To illustrate how fixed effects are estimated, we will focus on the estimation of γ_{00} and γ_{01}.

Grand mean ignoring grouping:	Grand mean (of means) ignoring precision:	**Precision weighted average**:
$\hat{\gamma}_{00} = \frac{\sum_{i=1}^{N} y_{ij}}{N}$	$\hat{\gamma}_{00} = \frac{\sum_{j=1}^{J} \overline{y_{.j}}}{J}$	$\hat{\gamma}_{00} = \frac{\sum_{j=1}^{J} \Delta_j^{-1} \overline{y_{.j}}}{\sum_{j=1}^{J} \Delta_j^{-1}}$

Where: $\Delta_j^{-1} = \frac{1}{\tau_{00} + (\sigma^2 / n_j)}$

σ^2 / n_j is the variance of $\overline{y}_{.j}$ as an estimator of β_{oj}. Dividing σ^2 by n_j controls for the fact that some level-2 units have greater variability simply because they are larger. τ_{00} is the variance of the true means, β_{oj}, about the grand mean, γ_{00}. As total variance decreases, precision increases.

What happens as n_j gets bigger?

What happens when the sample sizes are equal?

$$\hat{\gamma}_{01} = \frac{\sum_{i=j}^{J} \Delta_j^{-1} (W_j - \overline{W}_.^*)(\overline{Y}_{.j} - \overline{Y}_{..}^*)}{\sum_{i=j}^{J} \Delta_j^{-1} (W_j - \overline{W}_.^*)^2}$$

Where "*" indicates a precision weighted average.

In sum, a generalized least squares technique (a weighted technique) is used to estimate all fixed effects. The basic idea is that it gives greater importance to estimates from groups with larger sample sizes.

Random Effects

Random level-1 coefficients are estimated via empirical Bayes estimation. There are two ways to estimate β_{0j} (and remember, because it is allowed to vary across groups, we now need to estimate it separately *for each group*):

1. Based on the level-1 model: $\beta_{0j} = \overline{y}_{.j} - r_{ij}$
2. Based on the level-2 model: $\beta_{0j} = \gamma_{00} + u_{0j}$

Using Bayesian reasoning, we should use both. More specifically, let's use whichever gives you the best or optimal estimate *for each group*. The underlying idea is that, in some groups, you have greater precision than in others (once again because of the sample size).

In groups with greater precision (groups with a bigger sample size), the estimate should be based more on the level-1 model. In groups with less precision (groups with a smaller sample size), the estimate should be based more on the level-2 model. In groups with greater precision, we can say that the estimated group mean is a *reliable* estimate of the true group mean.

The optimal combination of 1 and 2 (the empirical Bayes estimator) is given by the equation:

$$\beta_{0j}^{*} = \lambda_j \bar{y}_{.j} + (1-\lambda_j)\gamma_{00}$$

Reliability is represented by: λ_j

$$\lambda j = \frac{\tau_{00}}{\tau_{00} + (\sigma 2 / n_j)}$$

$$\beta_{1j}^{*} = \lambda_j \beta_{1j} + (1-\lambda_j)\gamma_{10}$$

Variance/Covariance Components

These – $var(r_{ij})=\sigma^2$, $var(u_{0j})=\tau_{00}$, $var(u_{1j})=\tau_{11}$, $cov(u_{0j}, u_{1j})=\tau_{01}$ – are estimated via maximum likelihood (full and restricted ML are possible).

HYPOTHESIS TESTING

Available hypothesis tests (assuming *restricted maximum likelihood* estimation):

Single parameter hypothesis tests	Statistic
Does the average level of prejudice vary across countries?	χ^2
Does education affect prejudice?	t
Is prejudice higher in countries with a larger foreign born population?	t
Does the effect of education on prejudice vary across countries?	χ^2
Is the effect of education on prejudice stronger or weaker in countries with a larger foreign born population?	t

Multi parameter hypothesis tests	Statistic
Is the fit of my model better when I allow the effects of education, sex, and age to vary across countries compared to a model in which all three effects are fixed?	Likelihood ratio test (χ^2)

The goal of maximum likelihood estimation is to come up with the best estimate for some population parameter (e.g., σ^2 or τ_{00}). It uses the observed data and probability theory to find the most likely/probable population value given the sample data. In other words, the maximum likelihood estimate is the estimate that is most probable given our observed data. All maximum likelihood estimation is done iteratively – estimates are generated (e.g., for σ^2 and τ_{00}) and the probability for those estimates is then calculated. This is done over and over until the probability is maximized (when the 'likelihood function' is maximized).

The likelihood function can be used to evaluate the overall fit of the model. The 'deviance' is derived from the likelihood function – the deviance ranges from zero (indicating perfect fit) to positive infinity (indicating poor fit).

The deviance statistic can be used to answer the question listed above under multi parameter hypothesis tests. To answer the question, you would need to run two models:
1. A one way ANCOVA with random effects model controlling for education, sex, and age (all of which are fixed effects).
2. A random coefficient regression model controlling for education, sex, and age (all of which are random effects).

The *only* difference between the two models has to do with whether the slopes are random or fixed.

To conduct the likelihood ratio test, subtract the RCRM deviance from the ANCOVA deviance and test to see if the difference (which has a chi-square distribution) is significant. Remember that zero indicates a perfect fit, so if the fit of the model is better when you allow the effects to randomly vary across schools, then the RCRM deviance should be smaller.

For example (with hypothetical data):
H0: ANCOVA Deviance – RCRM Deviance=0
H1: ANCOVA Deviance – RCRM Deviance>0

ANCOVA Deviance=20,000; here the # of parameters to estimate=2, σ^2 and τ_{00}

RCRM Deviance=19,465; here the # of parameters to estimate=11, σ^2 and:

τ_{00}			
τ_{10}	τ_{11}		
τ_{20}	τ_{21}	τ_{22}	
τ_{30}	τ_{31}	τ_{32}	τ_{33}

20,000-19,465=535 (535 is your observed chi-square value)

The degrees of freedom for the chi-square test is 9 (11 parameters minus 2)

The critical chi-square value for 9 d.f. at $p<.05$ is 16.919

Therefore, you reject the null hypothesis (because the observed value is greater than the critical value) and conclude that you have significantly improved the fit of the model by allowing the three effects to randomly vary across schools. Note – this is a *joint* test!!!

What you can't do with restricted maximum likelihood estimation
One other common multi parameter hypothesis test has to do with whether or not the overall fit of your model is improved after controlling for a *set of variables*.

For example (Dependent variable=prejudice):
Model 1 IVs=sex and age
Model 2 IVs=sex, age, education, and social class

In this example, we might want to know whether the fit of the model is better after controlling for the SES variables. It is impossible to answer this question with restricted maximum likelihood estimation. One way of answering the question is to use *full* maximum likelihood estimation and the chi-square difference test described above.

MODEL BUILDING

Data analysis should always begin with a thorough examination of the univariate frequency distributions and descriptive statistics for each variable (to assess data quality, identify outliers, identify variables for transformation, etc.). Following this, I highly recommend exploratory bivariate analyses (e.g., plots to detect non-linearity, correlations, ANOVA, etc.) and multivariate analyses within each unit of analysis and between each unit of analysis. You should know your data before you begin more sophisticated analyses!!!

Building level-1 models

There are two general questions:

1. Should the variable be in the model?
2. If yes to question 1, should the effect be fixed, random, or non-randomly varying?

The best approach to model building is to use a "step-up" strategy – begin with a small set of theoretically relevant variables and fix their effects. Investigate the possibility of randomly varying effects for those with some theoretical basis. If the slope doesn't vary across groups, then fix it (also be aware of the reliability estimate, the number of iterations, etc. to help you decide)!

Above all else, use caution. Bryk and Raudenbush have found that they could only simultaneously estimate a maximum of 3 random slopes and the random intercept with data from 160 schools with an average school sample size (n_j) of 60. As the n_j goes down, it becomes more and more difficult to estimate randomly varying effects.

To delete a variable from the model, there should be:

1. No evidence of slope heterogeneity and
2. No evidence of average or fixed effects

Building level-2 models

Much of the previous discussion also applies to building level-2 models. The general rule of thumb for regression analysis is that you need 10 observations for each predictor variable.

If you want to predict a single level-2 outcome (e.g., a random intercept or a random slope), the number of observations is equal to the number of level-2 units and the general rule of thumb applies – e.g., if we had data for 30 countries and we wanted to predict differences in the intercept, then we could have 3 country-level independent variables.

The rough guidelines are not as clear when you have more than one level-2 outcome. B&R argue that the 10 observations rule is probably too liberal. If the level-2 outcomes are independent, then the 10-observation rule applies separately to each outcome.

In terms of model building, its best to build the model for the intercept first and then build models for the slope(s). I also strongly suggest that you begin with a small number of level-2 variables and slowly step them in to the model – e.g., begin with one variable and then step in a second. As you do this, examine changes in the slopes and changes in the standard errors.

WEIGHTS

Many publicly available datasets include one or more weights that should be applied in order to generalize from the sample to the population. These weights place greater emphasis on some cases compared to others in order to correct for differences in the probabilities of selection, errors in the sampling frame, or non-response. These are often referred to as sampling weights. They are referred to as 'pweights' in STATA.

It is possible to include weights at multiple levels of analysis in multilevel modeling – for example, at the person and country levels. Within the MIXED command, STATA allows pweights and fweights. Both of these are only available under full maximum likelihood estimation (and not restricted maximum likelihood estimation).

Some thoughts and cautions:

- Remember that group-specific sample sizes play an important role in estimation within a multilevel framework. They influence, for example, the precision, which is used to compute precision weighted averages of fixed effects. They also influence reliabilities, which are used to compute empirical Bayes estimates for random effects. What this means in practice depends upon your data?

    - People nested within occupations – My research suggests that more common occupations (based on data from the Bureau of Labor Statistics) tend to have more job incumbents in the General Social Survey. This is reassuring for those using the GSS given that the GSS is meant to be a probability sample. In a multilevel analysis with people nested within occupations, this means that occupations with more job incumbents will play a larger role in shaping the estimates of fixed and random effects. This seems acceptable to me.

    - People nested within countries – The countries included in cross-national data are not typically a probability sample of countries. Countries are included because researchers (and funding agencies) have decided to include them. Country-specific sample sizes may vary quite dramatically. Countries with larger samples will play a larger role in shaping the estimates of fixed and random effects. Imagine that the sample sizes for Germany and Czech Republic are 1,000 and 2,000, respectively. Is it desirable that the data from Czech Republic would play a larger role in shaping your estimates?

    - This issue may not be problematic when the level-2 units are countries because country-specific sample sizes are typically large. The precision will be high for all countries because of the large sample sizes (relevant for fixed effects). Empirical Bayes estimators are often referred to as shrinkage estimators. When the sample size for a group is small, its estimate (e.g., country mean or country slope) is shrunk toward the overall grand mean (intercept or slope). With large country-specific sample sizes, country reliabilities will be high and little shrinkage will occur.

    - Should we worry about differences in country-specific sample sizes?
        - It depends on what you are attempting to accomplish.
            - If you are trying to estimate the grand mean prejudice score or the grand mean education slope for 'Europe' then you may want to include a level-2 weight that makes the sample percentages equal to those in the population (see the table below and my STATA syntax).
            - If you are trying to estimate and predict group means/slopes (random effects), then this is less of an issue.

- Solutions?
  - Do nothing (see the final point below about including only a level-1 weight in STATA)
  - Design a level-2 weight to address the problem (see the table below and my STATA syntax)
  - Randomly select samples of the same size from the larger samples (this doesn't seem like a good option because you end up throwing data away)
  - Run the analyses using a variety of strategies and compare the results to see how robust they are to the differences in weighting method

	Population	%		Nj	%		Correction (% Pop/%sample)	Target Nj	%
slo	1,990,000	0.4		1,035	4.75		0.083376351	86	0.4
lv	2,516,000	0.5		1,031	4.73		0.105823502	109	0.5
irl	3,602,000	0.7		992	4.55		0.157457081	156	0.7
n	4,360,000	0.9		1,487	6.82		0.127146872	189	0.9
sk	5,332,000	1.1		1,388	6.37		0.16658306	231	1.1
a	8,047,000	1.6		1,007	4.62		0.346525094	349	1.6
bg	8,400,000	1.7		1,099	5.04		0.331445215	364	1.7
s	8,831,000	1.8		1,274	5.85		0.300587292	383	1.8
h	10,230,000	2.0		992	4.55		0.447192098	444	2.0
cz	10,331,000	2.1		1,106	5.08		0.405058168	448	2.1
nl	15,460,000	3.1		2,058	9.44		0.325757391	670	3.1
pl	38,587,600	7.7		1,568	7.20		1.067165727	1,673	7.7
e	39,210,000	7.8		1,221	5.60		1.392551732	1,700	7.8
i	57,204,000	11.4		1,091	5.01		2.273692907	2,481	11.4
gb	58,606,000	11.7		1,027	4.71		2.474581699	2,541	11.7
d	81,642,000	16.2		1,829	8.39		1.935664518	3,540	16.2
rus	148,140,992	29.5		1,585	7.27		4.052995686	6,424	29.5
	502,489,592	100.0		21,790	100.00			21,790	100.0

- You should NOT combine level-1 and level 2 weights into a single weight for use at level-1 in STATA

- If you include only a level-1 weight, STATA assumes that level-2 units are sampled with equal probability. This seems acceptable to me when conducting cross-national research.

STATA syntax:

```
mixed pbw [pweight=V342] || cntryid: , pwscale(size)
```

This syntax (above) includes a sampling weight at level 1 ('V342'). The 'pwscale(size)' option "specifies that first-level (observation-level) weights be scaled so that they sum to the sample size of their corresponding second-level cluster. Second-level sampling weights are left unchanged" (from the STATA manual). The average weight from the 2003 ISSP (V342) is not 1, so it is important to use the scaling option.

If you wanted to include a level-2 weight:

```
mixed pbw [pweight=V342] || cntryid: , pweight (l2weight) pwscale(size)
```

EXTENDED EXAMPLE
From: Kunovich, Robert M. 2004. “Social Structural Position and Prejudice: An Exploration of Cross-national Differences in Regression Slopes.” *Social Science Research*: 33, 1 (March): 20-44.

Variables
pbw is an 8 item scale (in z scores) measuring anti-immigrant prejudice
malem - female=0 male=1
agem - age measured in years
educm2 - education is measured in years
EGP=Erikson, Goldthorpe, and Portocarero Nominal Class Categories
EGP123 (reference category) - higher service, lower service, routine clerical and sales
EGP45 - independent and small employers
EGP711 - manual foremen, skilled manual, semi-unskilled manual, farm workers, farmers, farm managers
EGP21 - students
EGP22 - unemployed
EGP2325 - homemakers, retirees, and others not in the labor force
cntryid – Country id variable
WEUROPE – a dummy variable at level-2
LTIRMA5 – the five-year moving average long-term immigration rate

1. One-Way ANOVA with Random Effects: Does the level of prejudice vary across countries?

```
mixed pbw || cntryid:
estat group
estat icc
```

```
Mixed-effects ML regression                     Number of obs      =     21790
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       992
                                                               avg =    1281.8
                                                               max =      2058


                                                Wald chi2(0)       =         .
Log likelihood =  -28711.99                     Prob > chi2        =         .
```

pbw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	.1001793	.0906525	1.11	0.269	-.0774964	.277855

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cntryid: Identity				
var(_cons)	.1390369	.0479464	.0707288	.2733153
var(Residual)	.8132404	.0077943	.7981065	.8286612

```
LR test vs. linear regression: chibar2(01) =  2927.84 Prob >= chibar2 = 0.0000

. estat group
```

Group Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
cntryid	17	992	1281.8	2058

```
. estat icc

Intraclass correlation
```

Level	ICC	Std. Err.	[95% Conf.	Interval]
cntryid	.1460046	.0430148	.0799933	.2515926

$$ICC = \rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \frac{.1390369}{.1390369 + .8132404} = .1460046$$

STATA does not provide an estimate of the reliability (λ_j), but it can be calculated:

```
sort cntryid
by cntryid: egen nj=count(cntryid)
fre cntryid nj
```

cntryid

		Freq.	Percent	Valid	Cum.
Valid	2 d	1829	8.39	8.39	8.39
	4 gb	1027	4.71	4.71	13.11
	7 a	1007	4.62	4.62	17.73
	8 h	992	4.55	4.55	22.28
	9 i	1091	5.01	5.01	27.29
	10 irl	992	4.55	4.55	31.84
	11 nl	2058	9.44	9.44	41.28
	12 n	1487	6.82	6.82	48.11
	13 s	1274	5.85	5.85	53.96
	14 cz	1106	5.08	5.08	59.03
	15 slo	1035	4.75	4.75	63.78
	16 pl	1568	7.20	7.20	70.98
	17 bg	1099	5.04	5.04	76.02
	18 rus	1585	7.27	7.27	83.30
	24 e	1221	5.60	5.60	88.90
	25 lv	1031	4.73	4.73	93.63
	26 sk	1388	6.37	6.37	100.00
	Total	21790	100.00	100.00	

nj

		Freq.	Percent	Valid	Cum.
Valid	992	1984	9.11	9.11	9.11
	1007	1007	4.62	4.62	13.73
	1027	1027	4.71	4.71	18.44
	1031	1031	4.73	4.73	23.17
	1035	1035	4.75	4.75	27.92
	1091	1091	5.01	5.01	32.93
	1099	1099	5.04	5.04	37.97
	1106	1106	5.08	5.08	43.05
	1221	1221	5.60	5.60	48.65
	1274	1274	5.85	5.85	54.50
	1388	1388	6.37	6.37	60.87
	1487	1487	6.82	6.82	67.69
	1568	1568	7.20	7.20	74.89
	1585	1585	7.27	7.27	82.16
	1829	1829	8.39	8.39	90.56
	2058	2058	9.44	9.44	100.00
	Total	21790	100.00	100.00	

```
generate lambda_j = .1390369 / (.1390369 + (.8132404/nj))
tabstat lambda_j, statistics (mean sd) by (cntryid)
```

Note that σ^2 is assumed to be homogenous across countries. This assumption can be tested and relaxed if necessary.

```
Summary for variables: lambda_j
     by categories of: cntryid

 cntryid |      mean        sd
---------+--------------------
       d |  .9968122         0
      gb |  .9943369         0
       a |  .9942251         0
       h |  .9941383         0
       i |  .9946674         0
     irl |  .9941383         0
      nl |  .9971659         0
       n |  .9960819         0
       s |  .9954299         0
      cz |  .9947393         0
     slo |  .9943805         0
      pl |  .9962836         0
      bg |   .994706         0
     rus |  .9963233         0
       e |  .9952324         0
      lv |  .9943588         0
      sk |  .9958037         0
---------+--------------------
```

Sum=	16.919
Lambda=	0.995

The sum of the reliabilities for each country is 16.919. If you divide that sum by the number of countries (17), you get the reliability coefficient, which is 0.995. The reliability estimate of .995 suggests that the country sample means are quite reliable estimates of the true country population means (not surprising because the country sample sizes are large).

```
* Empirical Bayes Estimates of Country Means (Prejudice)
predict eb, reffects
sort cntryid
format eb %8.3f
tabstat eb, statistics (mean sd) by (cntryid)
```

cntryid	Empirical Bayes Estimates	Group Means	Country N
d	-0.191	-0.091	1829
gb	-0.104	-0.005	1027
a	-0.227	-0.128	1007
h	0.660	0.765	992
i	0.185	0.286	1091
irl	-0.898	-0.803	992
nl	-0.289	-0.189	2058
n	-0.072	0.028	1487
s	-0.276	-0.177	1274
cz	0.364	0.466	1106
slo	0.265	0.367	1035
pl	-0.086	0.014	1568
bg	0.346	0.448	1099
rus	0.031	0.131	1585
e	-0.452	-0.354	1221
lv	0.354	0.456	1031
sk	0.389	0.491	1388

2. One-way ANCOVA with Random Effects: What individual-level characteristics are associated with prejudice?

```
* Create grand mean centered variables
egen agegm=mean(agem)
fre agegm
generate agegrandc=agem-agegm
tabstat agem agegrandc, statistics( mean sd )

egen educgm=mean(EDUCM2)
fre educgm
generate educgrandc=EDUCM2-educgm
tabstat EDUCM2 educgrandc, statistics( mean sd )

mixed pbw malem agegrandc educgrandc EGP45 EGP711 EGP21 EGP22 EGP2325 || cntryid:
```

```
Mixed-effects ML regression                     Number of obs      =     21158
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       979
                                                               avg =    1244.6
                                                               max =      2025


                                                Wald chi2(8)       =   1047.74
Log likelihood = -27350.741                     Prob > chi2        =    0.0000
```

pbw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
malem	.044299	.0127677	3.47	0.001	.0192748	.0693231
agegrandc	.0017174	.0004935	3.48	0.001	.0007502	.0026846
educgrandc	-.0371156	.0019562	-18.97	0.000	-.0409496	-.0332816
EGP45	.064144	.0304606	2.11	0.035	.0044423	.1238456
EGP711	.1662578	.019623	8.47	0.000	.1277974	.2047183
EGP21	-.1559093	.0287295	-5.43	0.000	-.2122182	-.0996004
EGP22	.1077723	.0270251	3.99	0.000	.0548041	.1607405
EGP2325	.1431238	.019414	7.37	0.000	.1050732	.1811745
_cons	-.0101936	.0907542	-0.11	0.911	-.1880686	.1676813

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cntryid: Identity				
var(_cons)	.1366787	.0471309	.0695315	.2686709
var(Residual)	.7735129	.0075235	.7589068	.7884001

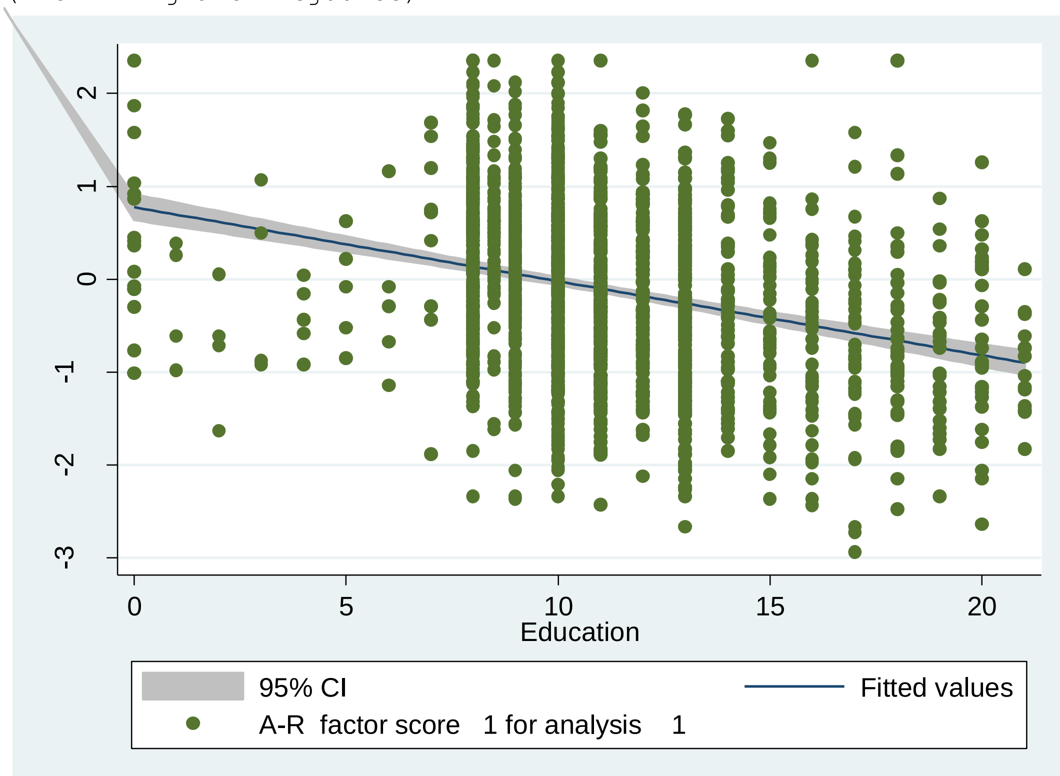
```
LR test vs. linear regression: chibar2(01) =  2920.64 Prob >= chibar2 = 0.0000
```

You can compute the percentage of explained variation at both levels by comparing the variance estimates across models. At the person-level: 4.9%.

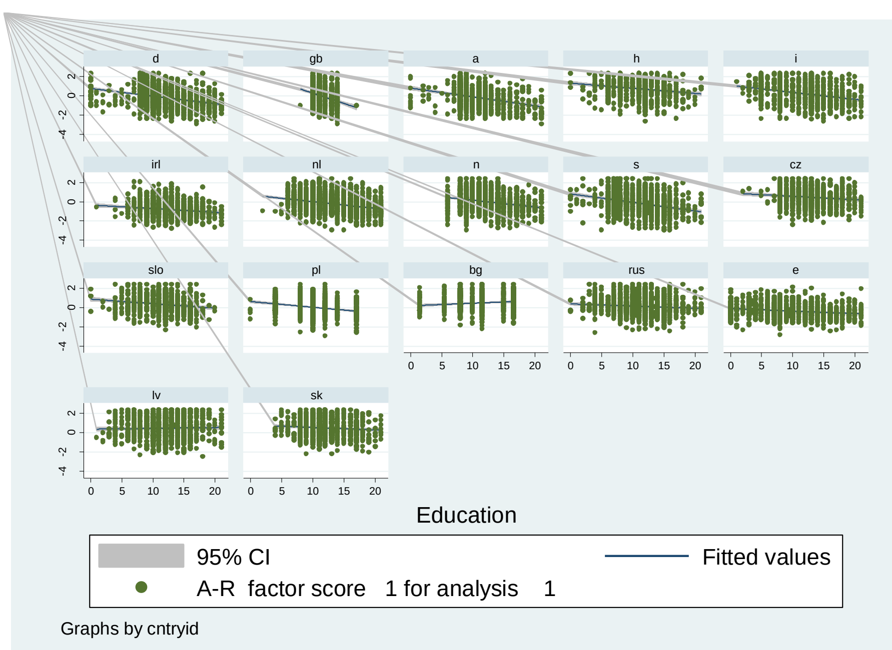
Notice that the variance component (tau) has been reduced from .1390369 to .1366787 (i.e., by about 1.7%). This suggests that differences in the average levels of the independent variables explain only about 1.7% of the country differences in prejudice. In other words, there is little evidence of composition effects here. Notice also that the country differences in prejudice remain significant.

3. Random Coefficient Regression Model: Does the relationship between prejudice and education vary across countries?

```
* You could start with scatterplots within each country:
twoway (lfitci pbw EDUCM2) (scatter pbw EDUCM2) if cntryid==2, xtitle(Education) ytitle
(Anti-immigrant Prejudice)
```




```
* You could create a trellis graph:
twoway (lfitci pbw EDUCM2) (scatter pbw EDUCM2), by (cntryid) xtitle(Education) ytitle
(Anti-immigrant Prejudice)
```



```
* You could run country-specific regressions (OLS).  Here are results for Germany:
bysort cntryid: regress pbw EDUCM2
```

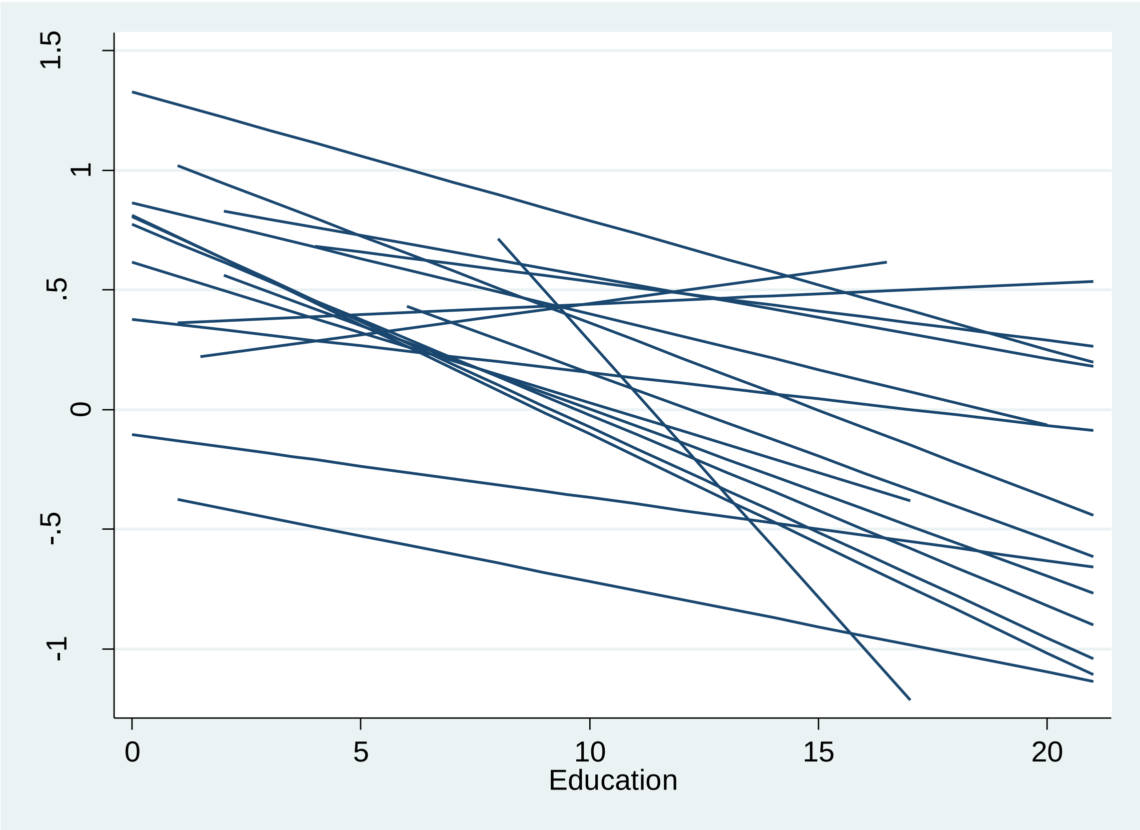
```
-> cntryid = d

      Source |       SS       df       MS              Number of obs =    1820
-------------+------------------------------           F(  1,  1818) =  165.02
       Model |  133.640874     1  133.640874           Prob > F      =  0.0000
    Residual |  1472.33192  1818  .809863544           R-squared     =  0.0832
-------------+------------------------------           Adj R-squared =  0.0827
       Total |   1605.9728  1819  .882887739           Root MSE      =  .89992

------------------------------------------------------------------------------
         pbw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      EDUCM2 |  -.0797096   .0062051   -12.85   0.000    -.0918795   -.0675398
       _cons |   .7749593   .0707867    10.95   0.000     .6361274    .9137911
------------------------------------------------------------------------------
```

```
* A spaghetti plot showing the education slopes for all countries:
statsby intere=_b[_cons] slopee=_b[EDUCM2], by (cntryid) saving(ols_educ): regress pbw
EDUCM2
sort cntryid
merge m:1 cntryid using ols_educ
drop _merge

generate prede = intere + slopee*EDUCM2
sort cntryid EDUCM2
twoway (line prede EDUCM2, connect(ascending)), xtitle(Education) ytitle(Fitted
regression lines)
```



```
* Group mean center education
egen educgrpm = mean(EDUCM2), by (cntryid)
generate educgroupc=EDUCM2-educgrpm

mixed pbw educgroupc || cntryid:
estimates store rie
mixed pbw educgroupc || cntryid: educgroupc, cov(unstructured)
estimates store rce
```

With education treated as fixed:

```
Mixed-effects ML regression                     Number of obs      =     21741
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       992
                                                               avg =    1278.9
                                                               max =      2058


                                                Wald chi2(1)       =    788.15
Log likelihood = -28256.829                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         pbw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  educgroupc |  -.0481391   .0017147   -28.07   0.000    -.0514999   -.0447783
       _cons |    .099866    .090695     1.10   0.271     -.077893     .277625
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntryid: Identity            |
                  var(_cons) |   .1391902   .0479901      .0708159    .2735814
-----------------------------+------------------------------------------------
               var(Residual) |   .7845093   .0075274      .7698938    .7994023
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  3030.07 Prob >= chibar2 = 0.0000
```

With education treated as random:

```
Mixed-effects ML regression                     Number of obs      =     21741
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       992
                                                               avg =    1278.9
                                                               max =      2058


                                                Wald chi2(1)       =     23.00
Log likelihood =   -28073.4                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         pbw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  educgroupc |  -.0549543   .0114576    -4.80   0.000    -.0774107   -.0324978
       _cons |   .0998614   .0906932     1.10   0.271    -.0778939    .2776168
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntryid: Unstructured        |
               var(educg~pc) |   .0021631   .0008127      .0010358    .0045174
                  var(_cons) |   .1391971   .0479874      .0708245    .2735754
         cov(educg~pc,_cons) |   .0050645   .0044542     -.0036655    .0137945
-----------------------------+------------------------------------------------
               var(Residual) |   .7692012   .0073838      .7548645    .7838102
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =  3396.93   Prob > chi2 = 0.0000
```

```
lrtest rce rie
```

```
. lrtest rce rie

Likelihood-ratio test                                 LR chi2(2)  =    366.86
(Assumption: rie nested in rce)                       Prob > chi2 =    0.0000

Note: The reported degrees of freedom assumes the null hypothesis is not on the boundary of the parameter
      space.  If this is not true, then the reported test is conservative.
```

4. Intercepts as Outcomes: What country-level characteristics are associated with prejudice?

```
mixed pbw malem agegrandc educgrandc EGP45 EGP711 EGP21 EGP22 EGP2325 LTIRMA5 || cntryid:
mixed pbw malem agegrandc educgrandc EGP45 EGP711 EGP21 EGP22 EGP2325 weurope || cntryid:
mixed pbw malem agegrandc educgrandc EGP45 EGP711 EGP21 EGP22 EGP2325 LTIRMA5 weurope ||
cntryid:
```

```
Mixed-effects ML regression                     Number of obs      =     21158
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       979
                                                               avg =    1244.6
                                                               max =      2025


                                                Wald chi2(9)       =   1055.31
Log likelihood = -27347.654                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         pbw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       malem |   .0443315   .0127676     3.47   0.001     .0193074    .0693557
   agegrandc |   .0017221   .0004935     3.49   0.000     .0007549    .0026892
  educgrandc |  -.0371069   .0019561   -18.97   0.000    -.0409407   -.0332731
       EGP45 |   .0638147   .0304603     2.10   0.036     .0041136    .1235157
      EGP711 |   .1661137    .019623     8.47   0.000     .1276533     .204574
       EGP21 |  -.1559762   .0287293    -5.43   0.000    -.2122847   -.0996678
       EGP22 |   .1074627   .0270249     3.98   0.000     .0544948    .1604306
     EGP2325 |   .1428622   .0194137     7.36   0.000      .104812    .1809125
     LTIRMA5 |   -.483378   .1772201    -2.73   0.006     -.830723    -.136033
       _cons |    .219337     .11341     1.93   0.053    -.0029425    .4416165
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntryid: Identity            |
                   var(_cons) |    .094882   .0327791      .0482078    .1867457
-----------------------------+------------------------------------------------
                var(Residual) |   .7735127   .0075235      .7589066       .7884
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  2180.09 Prob >= chibar2 = 0.0000
```

```
Mixed-effects ML regression                     Number of obs      =     21158
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       979
                                                               avg =    1244.6
                                                               max =      2025


                                                Wald chi2(9)       =   1065.79
Log likelihood = -27344.664                     Prob > chi2        =    0.0000
```

pbw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
malem	.044392	.0127676	3.48	0.001	.0193679	.0694161
agegrandc	.0017144	.0004934	3.47	0.001	.0007472	.0026815
educgrandc	-.0371168	.0019559	-18.98	0.000	-.0409503	-.0332833
EGP45	.0645451	.0304595	2.12	0.034	.0048457	.1242445
EGP711	.1658339	.019623	8.45	0.000	.1273734	.2042944
EGP21	-.1557085	.0287291	-5.42	0.000	-.2120165	-.0994005
EGP22	.1071813	.0270245	3.97	0.000	.0542143	.1601483
EGP2325	.1430077	.0194128	7.37	0.000	.1049593	.1810561
weurope	-.5305135	.1259624	-4.21	0.000	-.7773952	-.2836319
_cons	.2707652	.0925384	2.93	0.003	.0893933	.4521371

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cntryid: Identity				
var(_cons)	.066534	.0230573	.0337334	.1312281
var(Residual)	.7735129	.0075235	.7589067	.7884001

```
LR test vs. linear regression: chibar2(01) =  1536.86 Prob >= chibar2 = 0.0000
```

Mixed-effects ML regression

Group variable: cntryid

Number of obs	=	21158
Number of groups	=	17
Obs per group: min	=	979
avg	=	1244.6
max	=	2025
Wald chi2(10)	=	1067.78
Prob > chi2	=	0.0000

Log likelihood = -27344.197

pbw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
malem	.0443988	.0127676	3.48	0.001	.0193747	.0694229
agegrandc	.0017173	.0004934	3.48	0.001	.0007501	.0026844
educgrandc	-.0371119	.0019559	-18.97	0.000	-.0409453	-.0332784
EGP45	.0643184	.0304604	2.11	0.035	.0046171	.1240197
EGP711	.1658054	.019623	8.45	0.000	.1273449	.2042658
EGP21	-.1557687	.0287291	-5.42	0.000	-.2120767	-.0994606
EGP22	.1070824	.0270245	3.96	0.000	.0541154	.1600494
EGP2325	.1428789	.0194131	7.36	0.000	.1048299	.1809279
LTIRMA5	-.1752406	.1789425	-0.98	0.327	-.5259615	.1754804
weurope	-.4429883	.1516787	-2.92	0.003	-.740273	-.1457035
_cons	.3076636	.0976589	3.15	0.002	.1162556	.4990715

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cntryid: Identity				
var(_cons)	.0629609	.0218257	.0319153	.1242063
var(Residual)	.7735127	.0075235	.7589066	.7883999

LR test vs. linear regression: chibar2(01) = 1488.60 Prob >= chibar2 = 0.0000

5. Is the relationship between prejudice and education different in Western Europe?

```
mixed pbw malem agegrandc EGP45 EGP711 EGP21 EGP22 EGP2325 c.educgroupc##i.weurope ||
cntryid: educgroupc, cov(unstructured)
```

```
Mixed-effects ML regression                     Number of obs      =     21158
Group variable: cntryid                         Number of groups   =        17

                                                Obs per group: min =       979
                                                               avg =    1244.6
                                                               max =      2025


                                                Wald chi2(10)      =    363.94
Log likelihood = -27155.876                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------------
                pbw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
              malem |   .0486488   .0126524     3.85   0.000     .0238505    .0734471
          agegrandc |   .0019245     .00049     3.93   0.000     .0009642    .0028849
              EGP45 |   .0683076   .0302283     2.26   0.024     .0090612    .1275541
             EGP711 |    .153494   .0195125     7.87   0.000     .1152503    .1917377
              EGP21 |  -.1555371   .0289451    -5.37   0.000    -.2122684   -.0988057
              EGP22 |    .105526   .0268376     3.93   0.000     .0529252    .1581267
            EGP2325 |   .1424766    .019283     7.39   0.000     .1046827    .1802705
         educgroupc |  -.0141348   .0124942    -1.13   0.258    -.0386229    .0103533
          1.weurope |  -.5367378   .1247853    -4.30   0.000    -.7813124   -.2921632
                    |
weurope#c.educgroupc |
                 1  |  -.0541795   .0171537    -3.16   0.002    -.0878001   -.0205589
                    |
              _cons |   .2767274   .0916896     3.02   0.003     .0970192    .4564357
------------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntryid: Unstructured        |
              var(educg~pc)  |   .0011791   .0004658      .0005436    .0025576
                 var(_cons)  |   .0652966   .0226328      .0331017    .1288044
        cov(educg~pc,_cons)  |  -.0029915   .0023147     -.0075282    .0015451
-----------------------------+------------------------------------------------
              var(Residual)  |   .7580137   .0073763      .7436934    .7726097
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(3) =  1704.34   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

APPENDICES

Review of notation conventions

Units of analysis — level-1 units, i, n_j (e.g., respondents)
level-2 units, j, J (e.g., countries)

Dependent variable — Y (only possible at level-1)

Independent variables — level-1: $X_1, X_2, X_3, \ldots, X_q$

level-2: $W_1, W_2, W_3, \ldots, W_q$

Random effects — level-1: r_{ij}
level-2: u_{0j}, u_{1j}

Variance/covariance — level-1: σ^2_j
level-2: $var(u_{0j}) = \tau_{00}$
$var(u_{1j}) = \tau_{11}$
$cov(u_{0j}, u_{1j}) = \tau_{01}$

Coefficients — level-1: β_{oj}, β_{1j}
level-2: γ_{00}, γ_{01}
γ_{10}, γ_{11}

Six sub-models

1. The one-way ANOVA with random effects model (a.k.a. FUM):

Level-1 model:

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

Level-2 model:

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

Combined model:

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

STATA Command (multilevel linear model):

```
mixed dv || level2id:
```

β_{0j} is random and γ_{00} is fixed.
This model is fully unconditional at levels 1 and 2 (i.e., there are no independent variables at either level).

$Var(u_{0j}) = \tau_{00}$
$Var(r_{ij}) = \sigma^2_j$

Intraclass correlation:
$\rho = \tau_{00} / (\tau_{00} + \sigma^2_j)$

This model is used mainly to test whether or not the dependent variable varies across level-2 units – e.g., do some countries have higher average levels of prejudice than others. It can also be used to generate a point estimate and confidence interval for the grand mean as well as estimates of reliability – e.g., how reliable is the sample mean for country j as an estimator for the true group mean for country j?

2. The means as outcomes model:
Level-1 model:
$Y_{ij} = \beta_{0j} + r_{ij}$
Level-2 model:
$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$
Combined model:
$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + u_{0j} + r_{ij}$

STATA Command (multilevel linear model):
```
mixed dv level2iv1 level2iv2 level2iv3 || level2id:
```

β_{0j} is random and γ_{00} and γ_{01} are fixed.

In the FUM, Var(u_{0j}) or τ_{00} represents the total between-group variation in the dependent variable. Now it represents the residual variation – or the remaining/unexplained variance in the dependent variable after controlling for W_j. The only difference between the FUM and this model is the addition of the level-2 variable. Thus, this model is now conditional at level-2, but still unconditional at level 1 – e.g., there are no individual-level variables. This model is useful only if you are not interested in level-1 effects (rarely the case). You can use this model to explain differences in the average level of the dependent variable across groups – for example, is prejudice higher in countries with higher rates of immigration?

3. One-way ANCOVA with random effects model:
Level-1 model:
$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{..}) + r_{ij}$
Level-2 models:
$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10}$
Combined model:
$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{..}) + u_{0j} + r_{ij}$

STATA Command (multilevel linear model):
```
mixed dv level1iv1 level1iv2 level1iv3 || level2id:
```

β_{0j} is random and β_{1j}, γ_{00}, and γ_{01} are fixed.
In the FUM, Var(u_{0j}) or τ_{00} represents the total between-group variation in the dependent variable and Var(r_{ij}) or $\sigma^2{}_j$ represents the total within-group variation in the dependent variable. Now τ_{00} and $\sigma^2{}_j$ represent the residual variation – or the remaining/unexplained variance in the dependent variable after controlling for X.

This model is conditional at level-1 and unconditional at level-2 – there are no group-level predictor variables. This model is usually used to identify the average effects of the independent variables – for example, what is the average effect of education on prejudice across all countries?

4. Random coefficient regression model:
Level-1 model:
$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij}$
Level 2 models:
$\beta_{0j} = \gamma_{00} + u_{0j}$
$\beta_{1j} = \gamma_{10} + u_{1j}$
Combined model:
$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{.j}) + u_{1j}(X_{ij} - \overline{X}_{.j}) + u_{0j} + r_{ij}$

STATA Command (multilevel linear model):

```
mixed dv level1iv1 level1iv2 level1iv3 || level2id: level1iv1

Where level1iv1 is group mean centered
```

β_{0j} and β_{1j} are random and γ_{00}, and γ_{10} are fixed.

The only difference between this model and the one-way ANCOVA with random effects model is the inclusion of the random effect (u_{1j}) in the slope's level-2 model. This allows the slope of β_{1j} to vary across level-2 groups. This model is conceptually equivalent to the FUM. The FUM provides a test of whether or not groups have different the average levels of the dependent variable.

This model provides a test of whether or not the effect of the independent variable is different across the level-2 groups – e.g., does the effect of education on prejudice vary across countries?

One word of caution – it becomes more and more difficult to model and explain variation in slopes as n_j decreases. Think of how unreliable the slope estimate would be for a group with only 5 cases. If most of your groups have few cases, then it is difficult to distinguish between sampling error and true variance.

Notice that the level-1 variable is group mean centered – this is required whenever you allow the slope(s) to vary.

5. Intercepts and slopes as outcomes (a.k.a. the general model or the fully conditional model):
Level-1 model:
$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij}$
Level-2 models:
$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$
$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$
Combined model:
$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}(X_{ij} - \overline{X}_{.j}) + \gamma_{11}W_j(X_{ij} - \overline{X}_{.j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij}$

STATA Command (multilevel linear model):

```
mixed dv level1iv1 level1iv2 level2iv3 level1iv1##level2iv3 || level2id:
level1iv1

Where level1iv1 is group mean centered
```

β_{0j} and β_{1j} are random and γ_{00}, γ_{01}, γ_{10}, and γ_{11} are fixed.

We are back to the full model. It is conditional at all levels – that is, we have independent variables at both levels.

This submodel seeks to explain differences in the effects of level-1 variables and differences in the intercepts across level-2 units – e.g., use country-level variables to explain why the effect of education on prejudice varies across countries and why some countries have higher average levels of prejudice than others.

6. Nonrandomly varying slopes model:
Level-1 model:
$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij}$

Level-2 models:
$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$
$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j$

Combined model:
$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}(X_{ij} - \overline{X}_{.j}) + \gamma_{11}W_j(X_{ij} - \overline{X}_{.j}) + u_{0j} + r_{ij}$

β_{0j} is random, β_{1j} is non-randomly varying, and γ_{00}, γ_{01}, γ_{10}, and γ_{11} are fixed.
You can drop the random component when you explain all of the variance. This is an example of a nonrandomly varying slope model…it is also possible to do this for the intercept. Why bother? If there is no longer any significant variation in the slope or intercept after controlling for level-2 variables, then you can save degrees of freedom by eliminating the random effect(s) from the model.

THREE LEVEL MODELS

Pure Hierarchies

A classic example of a three-level model is students nested within classes and classes nested within schools. This is an example of a pure hierarchy because a student can be nested in one and only one classroom and a classroom can be nested within one and only one school.

$Y_{ijk} = \pi_{0jk} + e_{ijk}$, π is the mean for classroom j in school k; the error describes how each student in the same classroom varies from the classroom mean

$\pi_{0jk} = \beta_{00k} + r_{0jk}$, β is the mean for school k, the error describes how each class in the same school differs from the school mean

$\beta_{00k} = \gamma_{000} + u_{00k}$, γ is the grand mean, the error describes how each school differs from the grand mean

σ^2 is the within class variance
τ_π is the within school variance
τ_β is the between school variance

Taken together, these represent 100% of the variance. You can calculate the proportion of variation that is within classrooms, between classrooms within schools, and between schools by dividing each variance component by the total variation.

Stata syntax:

```
mixed dv iv1 iv2 || level3id: || level2id: , options
```

Example (People nested within regions nested within countries):

You can see in the cross-tabulation below that this is a pure hierarchy. Each region falls within only one country:

```
tab cntryid region if cntryid < 10
```

cntryid	region 251	252	253	254	255	256	257	Total
d	69	31	153	9	345	106	86	1,829
gb	0	0	0	0	0	0	0	1,027
a	0	0	0	0	0	0	0	1,007
h	0	0	0	0	0	0	0	992
i	0	0	0	0	0	0	0	1,091
Total	69	31	153	9	345	106	86	5,946

cntryid	region 258	259	260	261	262	263	264	Total
d	165	220	22	79	93	55	186	1,829
gb	0	0	0	0	0	0	0	1,027
a	0	0	0	0	0	0	0	1,007
h	0	0	0	0	0	0	0	992
i	0	0	0	0	0	0	0	1,091
Total	165	220	22	79	93	55	186	5,946

cntryid	region 265	266	401	402	403	404	405	Total
d	98	112	0	0	0	0	0	1,829
gb	0	0	92	57	93	91	96	1,027
a	0	0	0	0	0	0	0	1,007
h	0	0	0	0	0	0	0	992
i	0	0	0	0	0	0	0	1,091
Total	98	112	92	57	93	91	96	5,946

cntryid	region 406	407	408	409	410	411	701	Total
d	0	0	0	0	0	0	0	1,829
gb	81	40	109	201	101	66	0	1,027
a	0	0	0	0	0	0	41	1,007
h	0	0	0	0	0	0	0	992
i	0	0	0	0	0	0	0	1,091
Total	81	40	109	201	101	66	41	5,946

```
mixed pbw || cntryid: || region:
```

One-Way ANOVA with Random Effects Model

Mixed-effects ML regression   Number of obs   =   20785

Group Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
cntryid	16	992	1299.1	2058
region	218	5	95.3	418

Wald chi2(0)   =   .
Log likelihood = -27392.242   Prob > chi2   =   .

pbw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	.1674166	.0750911	2.23	0.026	.0202407	.3145925

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cntryid: Identity				
var(_cons)	.0871085	.0317912	.0425996	.1781214
region: Identity				
var(_cons)	.0230974	.0034655	.0172128	.0309939
var(Residual)	.8049307	.0079346	.7895285	.8206334

LR test vs. linear regression: chi2(2) = 2301.67   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

One-Way ANCOVA with Random Effects Model

```
mixed pbw malem agegrandc educgrandc EGP45 EGP711 EGP21 EGP22 EGP2325 || cntryid:
|| region:
```

```
Mixed-effects ML regression                     Number of obs      =     20161

-------------------------------------------------------------
                |     No. of       Observations per Group
 Group Variable |     Groups    Minimum    Average    Maximum
----------------+--------------------------------------------
        cntryid |         16        978     1260.1       2025
         region |        218          5       92.5        412
-------------------------------------------------------------

                                                Wald chi2(8)       =   1001.92
Log likelihood = -26063.583                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         pbw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       malem |   .0492756   .0130247     3.78   0.000     .0237477    .0748036
   agegrandc |   .0015843   .0005073     3.12   0.002     .0005901    .0025785
  educgrandc |  -.0380908   .0020076   -18.97   0.000    -.0420257   -.0341559
       EGP45 |   .0607795     .03111     1.95   0.051     -.000195     .121754
      EGP711 |   .1628643    .020114     8.10   0.000     .1234416     .202287
       EGP21 |  -.1592563   .0290904    -5.47   0.000    -.2162725   -.1022401
       EGP22 |   .1103763   .0277769     3.97   0.000     .0559345     .164818
     EGP2325 |   .1396592   .0198973     7.02   0.000     .1006612    .1786572
       _cons |   .0527759   .0757709     0.70   0.486    -.0957323     .201284
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
cntryid: Identity            |
                  var(_cons) |   .0861945   .0314614       .042149    .1762675
-----------------------------+------------------------------------------------
region: Identity             |
                  var(_cons) |   .0224224   .0034146      .0166363    .0302208
-----------------------------+------------------------------------------------
               var(Residual) |   .7652046   .0076607      .7503363    .7803675
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =  2278.38   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

```
mixed pbw malem agegrandc educgrandc EGP45 EGP711 EGP21 EGP22 EGP2325 weurope ||
cntryid: || region:
```

Mixed-effects ML regression   Number of obs   =   20161

Group Variable	No. of Groups	Observations per Group Minimum	Average	Maximum
cntryid	16	978	1260.1	2025
region	218	5	92.5	412

Wald chi2(9) = 1021.41

Log likelihood = -26057.309   Prob > chi2 = 0.0000

pbw	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
malem	.0493779	.0130246	3.79	0.000	.0238502	.0749055
agegrandc	.0015806	.0005072	3.12	0.002	.0005865	.0025747
educgrandc	-.0380595	.0020072	-18.96	0.000	-.0419936	-.0341253
EGP45	.0612865	.0311079	1.97	0.049	.0003161	.1222569
EGP711	.1624365	.0201143	8.08	0.000	.1230132	.2018598
EGP21	-.1591454	.0290896	-5.47	0.000	-.21616	-.1021307
EGP22	.1096516	.0277757	3.95	0.000	.0552121	.1640911
EGP2325	.1396134	.019895	7.02	0.000	.1006199	.178607
weurope	-.4411772	.1005014	-4.39	0.000	-.6381563	-.2441981
_cons	.2724356	.0721848	3.77	0.000	.1309559	.4139152

Random-effects Parameters	Estimate	Std. Err.	[95% Conf.	Interval]
cntryid: Identity				
var(_cons)	.0373782	.0143515	.0176115	.0793305
region: Identity				
var(_cons)	.0224874	.0034272	.0166806	.0303157
var(Residual)	.7651952	.0076605	.7503272	.7803578

LR test vs. linear regression: chi2(2) = 1259.19   Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.

Cross-classified Models

Sometimes the cases at multiple levels do not exist in a pure hierarchy. One example is having individuals nested within neighborhoods and occupations (they are cross-classified between neighborhoods and occupations). This is not a pure hierarchy because, for example, all of the people working within one common occupation will not live within the same neighborhood.

Cross-classified models can become quite complex because neighborhood characteristics could impact intercepts and/or slopes, occupation characteristics could influence intercepts and/or slopes, and the interaction between neighborhoods and occupations could impact intercepts and slopes. Often, however, we do not have sufficient data to examine the interaction of higher level units. Imagine a cross-tabulation between neighborhood and occupation id variables at level 1 – there would be many cells with zero cases in the cross-tabulation. The characteristics of your data will influence what analyses are possible. STATA is capable of estimating these types of models, but they are very slow!

Stata syntax:
```
mixed dv iv1 iv2 || _all: R.id1 || id2: , options
```

The grouping variable with more cases should be id1

Example from our harmonized data:
```
tab T_COUNTRY T_SURVEY_NAME
```

COUNTRY (TERRITORY) NAME	SURVEY PROJECT NAME							
	AMB	ASES	CB	CDCEE	CNEP	EB	EQLS	Total
AD	0	0	0	0	0	0	0	1,003
AL	0	0	0	0	0	0	0	5,588
AT	0	0	0	0	0	4,023	3,082	27,582
AZ	0	0	7,106	0	0	0	0	12,615
BA	0	0	0	0	0	0	0	3,599
BA-FBH	0	0	0	0	0	0	0	1,600
BA-RSR	0	0	0	0	0	0	0	800
BE	0	0	0	0	0	5,150	3,028	25,199
BE-FLA	0	0	0	0	0	0	0	7,385
BE-WAL	0	0	0	0	0	0	0	1,873
BG	0	0	0	2,095	0	3,025	3,037	34,384
BY	0	0	0	1,000	0	0	0	8,607
CH	0	0	0	0	0	0	0	27,616
CZ	0	0	0	1,683	0	3,143	3,234	43,822
DE	0	1,025	0	0	0	0	6,115	25,627

REFERENCES

* Enders, C.K. and D. Tofighi. 2007. "Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue." *Psychological Methods* 12:121-38.

Goldstein, Harvey. 1998. *Multilevel Statistical Models*. Chichester: John Wiley & Sons, LTD.

Heck, Ronald and Scott Thomas. 2000. *An Introduction to Multilevel Modeling Techniques*. Mahwah, N.J.: Lawrence Erlbaum Associates (Series: Quantitative Methodology Series, Methodology for Business and Management).

Hox, J.J. 1995. *Applied Multilevel Analysis*. Amsterdam: T.T. Publicaties.

* Hox, Joop. 2002. *Multilevel Analysis: Techniques and Applications*. Mahwah: Lawrence Erlbaum Associates, Inc.

* Kreft, Ita and Jan de Leeuw. 1998. *Introducing Multilevel Modeling*. London: Sage (Series: Introducing Statistical Methods).

Leyland, A. H. and H. Goldstein (Editors). 2001. *Multilevel Modelling of Health Statistics*. Chichester: John Wiley & Sons, LTD.

Luke, Douglas A. 2004. *Multilevel Modeling*. Thousand Oaks: Sage.

* Raudenbush, Stephen and Anthony Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edition). Thousand Oaks: Sage (Series: Advanced Quantitative Techniques in the Social Sciences Series).

Rabe-Hesketh, Sophia and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata, Volume I: Continuous Responses* (Third Edition). College Station, TX: Stata Press.

Reise, Steven P. and Naihua Duan (Editors). 2002. *Multilevel Modeling: Methodological Advances, Issues, and Applications*. Mahwah: Lawrence Erlbaum Associates, Inc.

* Snijders, Tom A.B. and Roel J. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications.

Note – I relied heavily on Raudenbush and Bryk (2002) to prepare this handout.