Multilevel Modeling (MLM)

part 1

Robert Yu

a few words before the talk

- *This is a report from attending a 2-day training course of* **Multilevel Modeling** *by Dr. Raykov Tenko, held on March 22 – 23, 2012, in Philadelphia, PA*
- *The original course was a commercial product of* Statistical Horizon Inc. *and it is strictly copyrighted on its content and handouts. We understand that no replication of it is allowed before permissions.*
- *This talk is intended to be an academic discussion of the topic and a report of this trip.*
- *We present the ideas, concepts and illustrations of the multilevel modeling technique fully according to the original lecturer, i.e. Dr. Tenko, of this course.*
- *To avoid simple replication of the materials, we paraphrase or reorganize the contents. These may be different from the original but representing our limited understanding of the subject.*
- *We don't have any business-intention or profit interest in this report talk. And we strongly prohibit anyone accessing this talk to do so.*

Original Course Plans
day 1

1. Why do we need multilevel modeling (MLM), and how come aggregation and disaggregation do not do the job?
2. The beginnings of MLM – Why what we already know about regression analysis is so useful, and centering of predictor variables.
3. The intra-class correlation coefficient – The underpinnings of a multilevel model.
4. How many levels? – Proportion third-level variance and how to evaluate it. Random intercept models, and model adequacy assessment.
5. Robust modeling of lower-level variable relationships in the presence of clustering effect.
6. Limitations and conclusion (Part 1).

Original Course Plans day 2

1. **What are mixed models, what are they made of, and why are they useful?**
2. **Random regression models – a general class of mixed/multilevel models of great utility**
3. **Mixed models with discrete response variables**
4. **Limitations and conclusion (Part 2). Outlook.**

Original: <u>Goals of workshop</u>

It is application oriented but with coherent discussion of theoretical issues, at an introductory/ intermediate level, with some more advanced issues.

Original: <u>Literature</u>

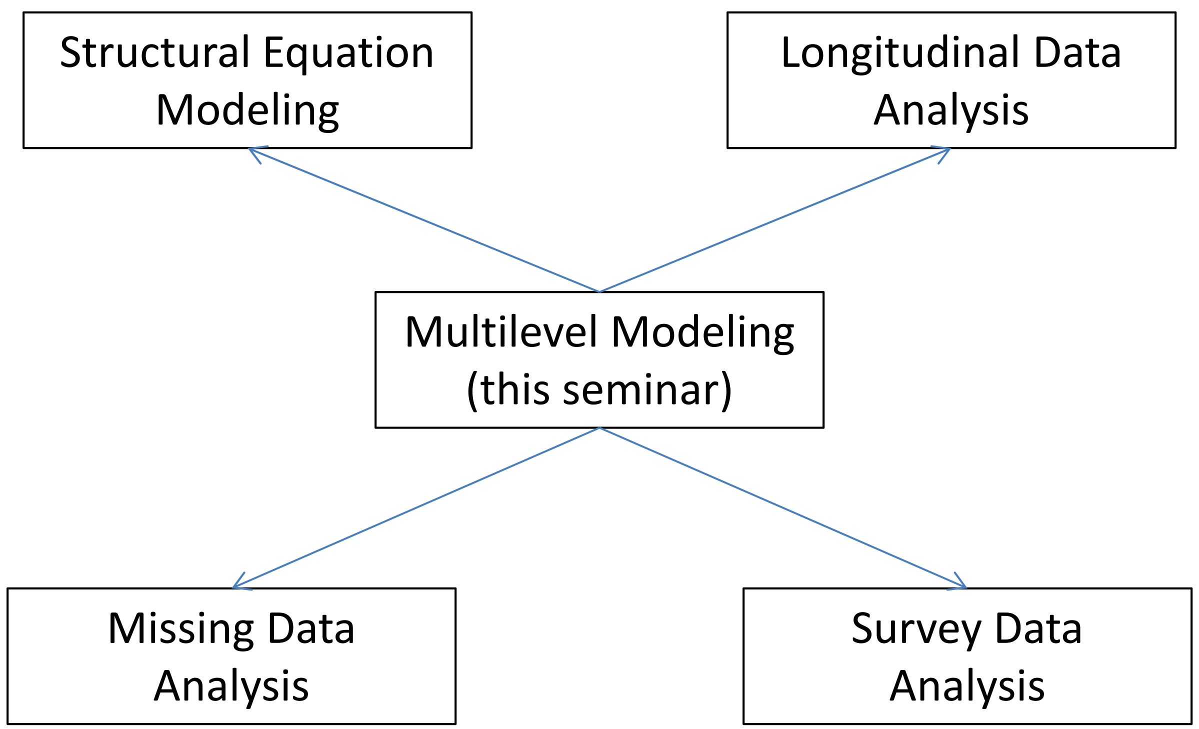
- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling with Stata.* College Station, TX: Stata Press.
- Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear and nonlinear modeling.* Thougsand Oaks, CA: Sage.
- Raykov, T. (2011). *Lecture notes on multilevel modeling.* Michigan State University.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent linear and mixed models*. Boca Raton, FL: Chapman & Hall.
- Snijders, T. A. B., & Bosker, R. (2012). *Multilevel models*. *An intro to basic an advanced modeling.* Thousand Oaks, CA: Sage.

Original: <u>Software</u>

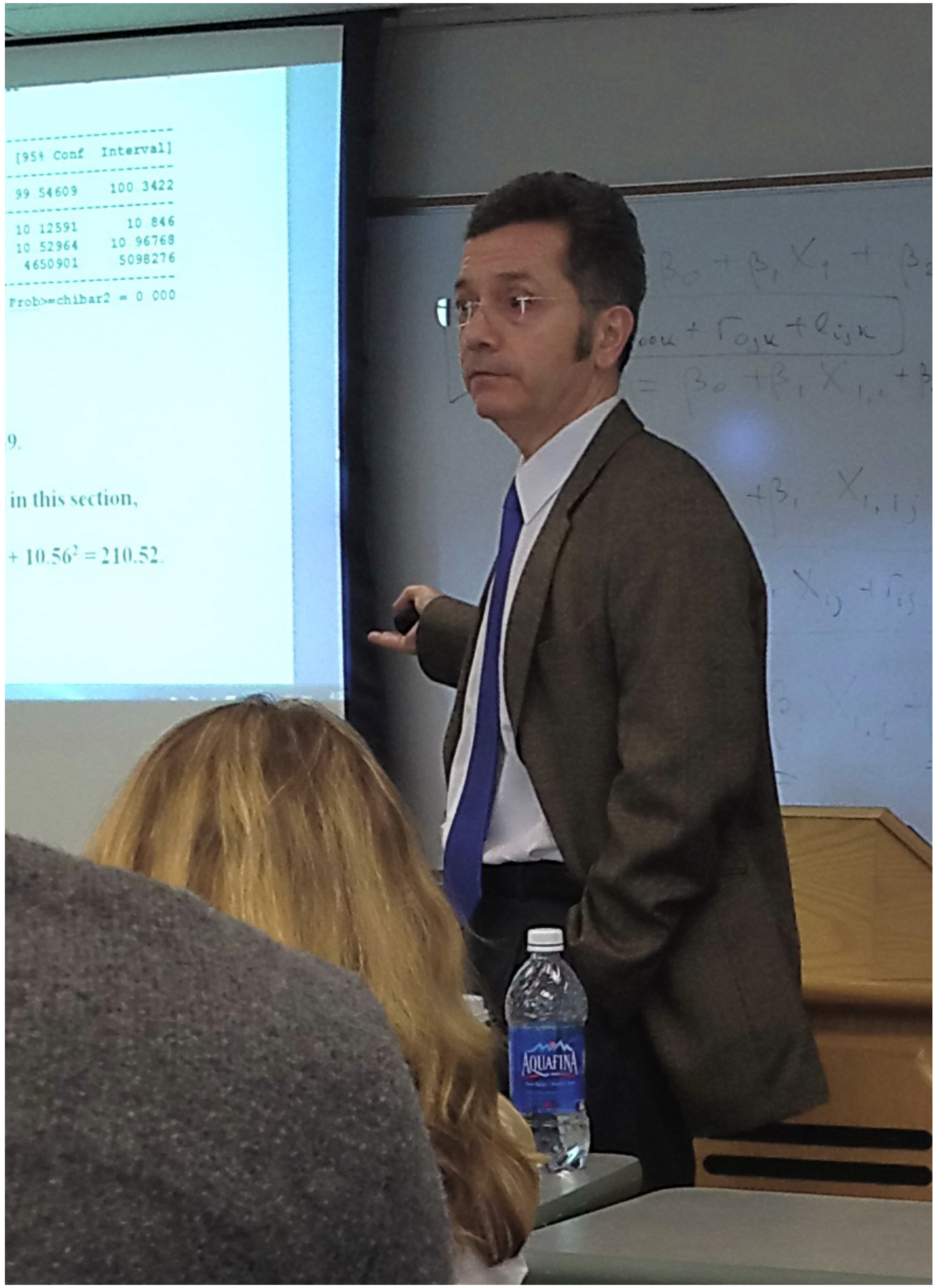
- **STATA**

Original note

This workshop provides a connection to the following main applied statistics areas (methodologies).




In the same way - and see Figure 1 below - we also come up with the interpretation of the slope as the average change in the dependent variable (DV), Y, which is associated with a unit change in the independent variable (IV), X. More specifically, in this example β1 is the mean increase in Y (MS) that is associated with an increase in X (IQ) by 1 unit.
That is, if we consider two groups of employees that differ in IQ by 1 unit (with IQ being constant within each of the groups), then their mean (expected) difference in MS is β1, according to the model.
Figure 1 Simple linear regression and parameter interpretation.


[95% Conf. Interval]
99.54609 100.3422
10.12591 10.846
10.52964 10.96768
4650901 5098276
Prob>=chibar2 = 0.000
in this section,
+ 10.56² = 210.52.
AQUAFINA


PSFS

Outlines for This Talk

1. **Motivation for MLM**
2. **Basics of MLM**
3. **ICC – decision to level**
4. **PTLV – how many levels**
5. **RMAC**
6. **Summary**

Listing of Stata Commands

1. Motivations for MLM

1. Motivations for MLM

- Data from studies in social, behavioral, biomedical sciences, and business exhibits distinct *hierarchical structure (multi-level)*.
- Because the studied units – individuals are *grouped* into larger units.
- The nesting may have an effect upon the subjects' outcome scores, some degree of correlation within the upper level.
- Examples
  - ❖ Employees are nested (clustered) within companies; former are level-1 units and the latter level-2 units.
  - ❖ Workers are nested within managers; employees are nested within teams. Workers/employees are level-1 units, managers/teams are level-2 units.
- The nesting implies a possibly serious *lack of independence* of individual scores, X, on the dependent variable(s), Y, of concern. The Y's within level-2 units are correlated, Y's are different among level-2 units.

1. Motivations for MLM - example

Employee (ID1)	Firm ID (ID2)	JS Score
1	Company 1	45
2	Company 1	46
3	Company 1	44
4	Company 1	42
5	Company 2	79
6	Company 2	78
7	Company 2	77
8	Company 2	75
9	Company 3	92
10	Company 3	91
11	Company 3	93
12	Company 3	94
…	…	…
50	Company 30	66

Note: JS = Job satisfaction score

The individual worker's JS *scores within company seem to be relatively similar (correlated).*

1. Motivations for MLM - example

- Another type of examples is when *multiple measures* obtained from each subject, which can be viewed as *nested within subject*. Longitudinal studies are typical cases.

- Thus, level-1 units are the repeated measurements, and subjects are the level-2 units.

1. Motivations for MLM

- Traditional methods, e.g. regression analysis, ANOVA, multivariate statistics, analysis of qualitative data, etc., are called *single-level methods*, and they assume that these subjects' scores are *independent.*
- To the degree to which this assumption is violated, the results of an application of those methods on hierarchical data will yield *less trustworthy if not even misleading* results.
- A frequent consequence of a serious violation of the above independence assumption is the phenomenon of *spurious significance*, if this violation is neglected.
- *Traditional Mistakes:* **aggregation** or **disaggregation** in dealing hierarchical data.
  - ❖ Aggregation - scores on the lowest level of the data hierarchy are summarized (usually aggregated) to scores that are then considered such of the higher-order unit they belong to, e.g. taking *means* within level-2 unit.
  - ❖ Disaggregation - to *disaggregate* the data down to the level-1 units and use classical (single-level) methods for analysis. In study of employees' sales ~ profit in a nationwide marketing company, disregarding regional difference, *flatting* all employees from their groups.

1. Motivations for MLM

- Errors with **aggregation**
  - ‘shift of meaning’
  - ‘ecological fallacy’
  - Precluding examining *cross-level interaction effects*
- Errors with **disaggregation**
  - *Destroyed* the interrelationships within a secondary level unit
  - *Exaggerates* the effect of sample size through flatting the secondary level
- Both *aggregation* and *disaggregation* lead to very different results. Typically, neither leads to the correct results.
- **Analytic benefits of multilevel modeling**
  - Improved estimation of effects within individual level-2 units

    By estimating regression coffients within a level-2 unit can *be strengthened* through finding similar estimates from other level-2 units.
  - Modeling cross-level effects

    In hierarchical data, relationships at level-1 are *moderated* by charateristics of level-2 units, e.g. the relationship between job satisfaction and quality of life of employees may be affected by the size of the company they work for. MLM helps properly estimate these effects.
  - Partitioning of Variance-Covariance Components

    Breaking down the variance of a response variable at a given level into within-level sources and such stemming from differences among higher-order units.

2. Basics of MLM

2. Basics of MLM

- MLM could be viewed as 'multiple multiple regression modeling', or as sets of regression equations.
- In studying satisfaction with company management (MS; *Y*) in an industry by employees' intelligence (IQ; *X*):

$$MS = \beta_0 + \beta_1\, IQ + \text{residual}, \quad \text{or,} \quad Y = \beta_0 + \beta_1 X + r$$

Company 1: $Y_1 = \beta_{0,1} + \beta_{1,1} X_1 + r_1$,

Company 2: $Y_2 = \beta_{0,2} + \beta_{1,2} X_2 + r_2$,

. . .

Company *J*: $Y_J = \beta_{0,J} + \beta_{1,J} X_J + r_J$.

$Y_j = \beta_{0j} + \beta_{1j} X_j + r_j$ $(j = 1, \ldots, J)$

where residual *r* and variance σ^2 : $r \sim N(0, \sigma^2)$.

- β_{0j} and β_{1j} are *random variables* (i.e. companies were randomly picked) and not observed, i.e. latent variables, with mean $E(\beta_0)=\gamma_{00}$, $E(\beta_1) = \gamma_{10}$, variance $Var(\beta_0)=\tau_{11}$, and covariance $Cov(\beta_0, \beta_1)= \tau_{01}$.
- MLM, including high-order level, *e.g. the 2-level model*

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \qquad (j = 1, \ldots, J,\ i = 1, \ldots, n_j)$$

e.g. *i – employees, 1st level, j – company, 2nd level*

2. Basics of MLM - example

use "C:\TEACH\Multilevel Modeling\Workshops\Misc\lea.dta", clear

d

```
Contains data from C:\TEACH\Multilevel Modeling\Workshops\Misc\lea.dta
  obs:           887
 vars:             3                          14 Jul 2011 16:23
 size:        14,192 (99.9% of memory free)
-----------------------------------------------------------------------
storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------
firm            float  %9.0g                  firm ID
js              float  %9.0g                  job satisfaction score
iq              float  %9.0g                  ingelligence test score
-----------------------------------------------------------------------
Sorted by:  firm  iq
```

sum

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        firm |       887    26.05637    14.12965          1         48
          js |       887    30.56595    6.666168          5         40
          iq |       887    100.3608    5.783583         79        111
```

2. Basics of MLM - example

graph twoway (scatter js iq, mlabel(firm)) (lfit js iq) if firm <=8

The line through the points would be the single-level regression line, $js_{ij} = b_0 + b_1\, iq_{ij}$

/* single-level method, *ignoring the firm effects*: */

reg js iq

```
      Source |       SS       df       MS              Number of obs =     887
-------------+------------------------------           F(  1,   885) =  341.40
       Model |  10960.2737     1  10960.2737           Prob > F      =  0.0000
    Residual |  28411.6181   885  32.1035233           R-squared     =  0.2784
-------------+------------------------------           Adj R-squared =  0.2776
       Total |  39371.8918   886  44.4378011           Root MSE      =   5.666

------------------------------------------------------------------------------
          js |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .6081306   .0329126    18.48   0.000     .5435347    .6727265
       _cons |   -30.4665    3.30861    -9.21   0.000    -36.96014   -23.97286
------------------------------------------------------------------------------
```

The single-level model result: **js = -30.52 + .61 × iq**

2. Basics of MLM - example

/* Multi-level method: */

xtmixed js iq || firm:, variance

```
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log likelihood = -2767.8923
Iteration 1:   log likelihood = -2767.8923
Computing standard errors:
Mixed-effects ML regression                     Number of obs      =       887
Group variable: firm                            Number of groups   =        48
                                                Obs per group: min =         5
                                                               avg =      18.5
                                                               max =        62
                                                Wald chi2(1)       =    347.92
Log likelihood = -2767.8923                     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
          js |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .6088066   .0326392    18.65   0.000     .5448349    .6727783
       _cons |  -30.51571   3.295415    -9.26   0.000    -36.9746   -24.05682
------------------------------------------------------------------------------

  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
firm: Identity               |
                  var(_cons) |   4.026853   1.189895      2.256545    7.186004
-----------------------------+------------------------------------------------
               var(Residual) |   28.12721    1.37289       25.5611    30.95094
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =    56.38 Prob >= chibar2 = 0.0000
```

The last line tests the hypothesis – given this model – that we can do better using this multilevel model than standard, single-level regression.

The MLM (2-level) modeling gives **(i) the degree of variability in the parallel regression lines of firm variability upon job satisfaction around this population line, var(_cons); as well as (ii) the error variance, var(Residual).**

2. Basics of MLM - example

/* to evaluate these firm effects, with standard errors (SE): */

```
predict re, reffects          /* evaluate effects */
predict se, reses             /* evaluate standard errors */
list in 1/40                  /* print to screen */
```

```
     +--------------------------------------+
     | firm   js    iq          re        se |
     |--------------------------------------|
  1. |    1   32    89   -2.676116   .9377579 |
  2. |    1   20    89   -2.676116   .9377579 |
  3. |    1   18    90   -2.676116   .9377579 |
  4. |    1   25    91   -2.676116   .9377579 |
  5. |    1   11    91   -2.676116   .9377579 |

              ... ...

 21. |    1   27   105   -2.676116   .9377579 |
 22. |    1   22   106   -2.676116   .9377579 |
 23. |    1   32   107   -2.676116   .9377579 |
 24. |    1   27   109   -2.676116   .9377579 |
 25. |    1   33   109   -2.676116   .9377579 |
     |--------------------------------------|
 26. |    2   19    89   -.0152072   1.286861 |
 27. |    2   35    90   -.0152072   1.286861 |
 28. |    2   27    95   -.0152072   1.286861 |
 29. |    2   34   100   -.0152072   1.286861 |
 30. |    2   37   100   -.0152072   1.286861 |
     |--------------------------------------|
 31. |    2   28   101   -.0152072   1.286861 |
 32. |    2   28   103   -.0152072   1.286861 |
 33. |    2   27   106   -.0152072   1.286861 |
 34. |    2   32   108   -.0152072   1.286861 |
 35. |    2   37   109   -.0152072   1.286861 |
     |--------------------------------------|
 36. |    3   37    96    1.058414   1.370049 |
 37. |    3   33    99    1.058414   1.370049 |
 38. |    3   22   100    1.058414   1.370049 |
 39. |    3   37   101    1.058414   1.370049 |
 40. |    3   32   104    1.058414   1.370049 |
     +--------------------------------------+
```

3. ICC – decision to level

3. ICC – decision to level

- With a hierarchical data set, 2 things to consider: (1) make sure the uncorrelatedness assumption in the data is violated – decision to have high-order, (2) how many levels are sustained by the data.
- ICC – Intraclass Correlation Coefficient, is a measurement of how much correlatedness in a hierarchical data set.
- In a 2-level MLM, using the conventional one-way ANOVA model,

$$Y_{ij} = \mu_j + e_{ij} \ (= \mu + \alpha_j + e_{ij}).$$

Y_{ij} is the dependent variable score for the ith subject in the jth group, μ_j is the mean of the latter on this variable, and e_{ij} is the associated residual, i.e., individual-specific deviation of his/her Y score from the corresponding group mean ($j = 1, \ldots, J$, $i = 1, \ldots, n_J$). This is an equation *only with an intercept*, the mean of the jth level-2 unit.

- In MLM, level-2 units are randomly picked, and thus μ_j (α_j) are random. Comparing with classical one-way ANOVA with fixed effects, this model

$$\boldsymbol{Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}} \qquad \boldsymbol{(j = 1, \ldots, J,\ i = 1, \ldots, n_j)}$$

is with random effect

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

where $Var(r_{ij}) = \sigma^2$, group mean $\boldsymbol{\beta_{0j}} = \gamma_{00} + \mu_{0j}$ *i.e. grand mean + a random quantity, assuming uncorrelated with* $\boldsymbol{r_{ij}}$.

3. ICC – decision to level

- Taking variance from both sides,

$$Var(Y_{ij}) = Var(\beta_{0j} + r_{ij}) = \tau_{00} + \sigma^2.$$

  where $\tau_{00} = Var(\beta_{0j})$.
- "**How much variance in the response variable *Y* stems from between-group differences (and how much from within-group differences)?**"

  Definition of ICC: $\rho = \tau_{00} / (\tau_{00} + \sigma^2)$.

  i.e. ICC = *percentage of between-group variability in observed variance*.

  Note, in Stata, ICC = $\Psi/(\Psi + \Theta)$.
- "**What is the relevance of the value of ICC?**"

  A high ICC implies "the between-group variance dominates the within-group variance. And thus, most of the differences that we see across subjects on *Y* are actually stemming from group differences." Conversely, a small ICC points to the variance in observed response stems from individual differences within groups (level-2 units).
- "**Why is the ICC called a correlation?**"

$$\begin{aligned} Corr(Y_{ij}, Y_{i'j}) &= Corr(\beta_{0j} + r_{ij}, \beta_{0j} + r_{i'j}) \\ &= Cov(\beta_{0j} + r_{ij}, \beta_{0j} + r_{i'j}) / [\boldsymbol{Var}(\beta_{0j} + r_{ij}), \boldsymbol{Var}(\beta_{0j} + r_{i'j})]^{1/2} \\ &= Var(\beta_{0j}) / Var(\beta_{0j} + r_{ij}) \\ &= \tau_{00} / (\tau_{00} + \sigma^2) = \rho . \end{aligned}$$

3. ICC – decision to level

- How to use ICC?
  - *"… no hard and fast rule that would tell us how large an ICC is large enough to proclaim non-negligible nesting (clustering, lack of independence) effect*. "
  - "… an ICC well into the double digits (on the percentage scale) could often be considered large enough to justifying a two-level analysis (i.e., application of a multilevel model as opposed to a classical statistical modeling approach). "
  - "… a simplified 'rule' submits that an ICC well into the single digits on the percentage scale may be viewed as suggestive of single-level analysis may still be appropriate (i.e., conventional regression analysis or another appropriate analysis ignoring the hierarchical nature of the data). "
- Observation: "the [*classical]* one-way random effect ANOVA … is also called a *two-level* (*fully*) *unconditional model*. The reason is that at none of its two levels we have included any predictor (explanatory variable, regressor, or covariate). "
- Confidence Interval (CI) of ICC can provide more info over point estimate ICC.
- CI of ICC can be obtained in Stata using command 'xtreg'.

3. ICC – decision to level - example

- JS ~ IQ data, n=7185 workers from J=160 companies.

```
     +-----------------------------------------+
     |   id   minority   female      iq      js |
     |-----------------------------------------|
  1. | 1224          0        1  -1.528   5.876 |
  2. | 1224          0        1   -.588  19.708 |
  3. | 1224          0        0   -.528  20.349 |

    ...  ...
 19. | 1224          0        1    .042   2.927 |
 20. | 1224          0        0   -.078  16.405 |
     +-----------------------------------------+
```

xtreg js, i(id) mle

```
Iteration 0:   log likelihood = -23558.083
Iteration 1:   log likelihood = -23557.906
Iteration 2:   log likelihood = -23557.905

Random-effects ML regression                    Number of obs      =      7185
Group variable: id                              Number of groups   =       160

Random effects u_i ~ Gaussian                   Obs per group: min =        14
                                                               avg =      44.9
                                                               max =        67
                                                Wald chi2(0)       =      0.00
Log likelihood  = -23557.905                    Prob > chi2        =         .
------------------------------------------------------------------------------
     js      |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   12.63707   .2436216    51.87   0.000     12.15958    13.11456
-------------+----------------------------------------------------------------
    /sigma_u |   2.924631   .1826925                      2.587612    3.305544
    /sigma_e |   6.256868   .0527937                      6.154245    6.361202
         rho |   .1793109   .0185934                      .1452078    .2180551
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)=  983.92 Prob>=chibar2 = 0.000
```

Note: in Stata, ρ as "rho", and σ^2 as 'sigma_u'.

3. ICC – decision to level - example

- The resulting 95% CI is *not symmetric* around the ICC estimae: 0.18 (0.14, 0.22).
- The CI may provide information about how close this ICC is to the population ICC.
- If a ‘threshold’ ICC ρ_0 has been established in a domain, one can ‘test’ the hypothesis “ICC > ρ_0“ in the population by using the above CI:
  - If ICC > ρ_0, retain that hypothesis.
  - If CI is entirely positioned below ρ_0, reject it.
- Model choice issue could be more complicated in a particular substantive domain and empirical setting.

4. PTLV – how many levels

4. PTLV – how many levels

- **Proportion Third-Level Variance** – PTLV
- Example of a 3-level model:

$$\boldsymbol{Y_{ijk} = \pi_{0jk} + e_{ijk}} \ ,$$
$$\boldsymbol{\pi_{0jk} = \beta_{00k} + r_{0jk}} \ ,$$
$$\boldsymbol{\beta_{00k} = \gamma_{000} + \mu_{00k}} \ .$$

- Interval estimation of the ratio of third-level variance to level-1 variance for a contemplated three-level model:

$$\delta = \frac{\tau_\beta}{\sigma^2 + \tau_\pi + \tau_\beta}.$$

- When this δ is “large enough”, a 3-level model is warranted.
- The δ ‘s CI may be indispensible in such a decision process.
- A 2-step procedure for point and interval estimation of PTLV but one single command in Stata: xtmixed
  - Step 1. to fit 3-level model and get the *`variance`* and *`convariances`*.
  - Step 2. to obtain PTLV δ and its 95% CI.

4. PTLV – how many levels

Example

```
Contains data from C:\T E A C H\Workshops\SH\MLM\Data\3L_income.dta
  obs:         3,153
 vars:            11                          20 Aug 2011 18:22
 size:       138,732
-------------------------------------------------------------------------------
variable name:  (float  %9.0g)
region, state, age, gender, marital, hours, citizen, person, constant, degree, income
-------------------------------------------------------------------------------
```

(Step 1) **xtmixed income||region:||state:**

```
Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log likelihood = -3987.4675  Iteration 1:   log likelihood = -3987.4675
Computing standard errors:
Mixed-effects ML regression                     Number of obs      =      3152
-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
         region |        9        139      350.2        548
          state |       51          5       61.8        221
-----------------------------------------------------------
                                                Wald chi2(0)       =         .
Log likelihood = -3987.4675                     Prob > chi2        =         .
------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   10.08074   .1582428    63.70   0.000     9.770594    10.39089
------------------------------------------------------------------------------


------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
region: Identity             |
                   sd(_cons) |   .4624112   .1155155      .2833927    .7545151
-----------------------------+------------------------------------------------
state: Identity              |
                   sd(_cons) |   .2087944   .0324538      .1539618    .2831554
-----------------------------+------------------------------------------------
                sd(Residual) |   .8439511    .010721      .8231977    .8652277
------------------------------------------------------------------------------
LR test vs. linear regression:      chi2(2) =   863.67   Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

4. PTLV – how many levels

- Explanation of this Stata command, `xtmixed`,
  - In the example, the levels go `person` => `state` => `region`.
  - Command `xtmixed` takes response variable (`income`) and then the highest leveled variable (`region`) – 3rd level, and then lower nesting leveled variable (`state`) – 2nd level.
- Check the covariance matrix of these 3 variances, i.e. the covariance matrix of the logarithms of the above standard deviation estimates (estimates of the square roots of the 3 variances, in $\delta = \frac{\tau_\beta}{\sigma^2+\tau_\pi+\tau_\beta}$).

```
mat list e(V)
      symmetric e(V)[4,4]
                        income:   lns1_1_1:   lns2_1_1:    lnsig_e:
                         _cons       _cons       _cons       _cons
  income:_cons       .02504077
lns1_1_1:_cons               0   .06240561
lns2_1_1:_cons               0  -.00096802   .02415984
 lnsig_e:_cons               0  -1.007e-06  -.00007521   .00016138
```

```
(Step 2) Transform them back by: to exponentiate and then square the estimated square-rooted
variances of interest, which is achieved in Stata as follows:
local var_r exp([lns1_1_1]_b[_cons])^2
local var_s exp([lns2_1_1]_b[_cons])^2
local var_e exp([lnsig_e]_b[_cons])^2
Using Stata command ‘nlcom’ to work out point and interval estimates.
nlcom (PTLV: `var_r'/(`var_e'+`var_s'+`var_r'))
PTLV:  exp([lns1_1_1]_b[_cons])^2/(exp([lnsig_e]_b[_cons])^2+exp([lns2_1_1]_b[_cons])^2+
exp([lns1_1_1]_b[_cons])^2)
------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        PTLV |   .2205116   .0861043     2.56   0.010     .0517503     .389273
------------------------------------------------------------------------------
```

Nearly a quarter of the obseved variability in income across workers is due to regional differences, and the significant 95% CI may conclude a 3-level modeling is needed.

4. PTLV – Random Intercept Models

- Random Intercept Models (RIMs) – multilevel models with covariates
- RIMs are the next step up from ANOVA Random Effects model.
- In a RIM, covariates are included but none of them is allowed to have a cluster-specific effect upon the response, i.e. *each covariate's effect is assumed to be the same in the clusters.* **Note:** in the 2nd part of the talk, *one or more of these covariates are allowed to have different effect across clusters.*
- Example: relationship between mastery test scores of employees and a host of explanatory variables including worker's gender and manager characteristics.

```
obs:          8,604
 vars:             10                           10 Aug 2011 16:52
 size:        378,576 (99.3% of memory free)
------------------------------------------------------------------------------------
variable name   type: float,  format: %9.0g
------------------------------------------------------------------------------------
manager, mage, meduc, male, married, hsgrad, somecoll, collgrad, black, masttest
------------------------------------------------------------------------------------
manager = manager ID;
masttest = worker's score on a professional mastery test;
mage = age;
male = gender (1 for male, 0 for female);
married = 0/1-variable for manager being married;
hsgrad = 0/1-variable for manager having HS diploma;
somecoll = 0/1-variable for manager (some college education, like community college);
collgrad = 0/1-variable for manager having college degree;
black = 0/1-variable for manager being black;
```

- Conventional regression model:

$$y_{ij} = \beta_0+\beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \ldots + \beta_p X_{p,ij} + e_{ij}, \quad (i=1,\ldots,n_j, j=1,\ldots,J)$$

$X_1, X_2, \ldots, X_p$ are covariates.

4. PTLV – Random Intercept Models

- In this modeling, covariates are fixed and only the error term is random.
- To accommodate the *clustering* effect of the managers, split the error term into
  (i) a 'common' part associated only with the manager
  (ii) an unrelated remaining contribution associated only with his/her employees

$$e_{ij} = \zeta_j + \varepsilon_{ij}$$

  Then the above becomes

$$\begin{aligned} y_{ij} &= \beta_0+\beta_1 X_{1,ij} +\beta_2 X_{2,ij} +\ldots+\beta_p X_{p,ij} +e_{ij} \\ &= \beta_0+\beta_1 X_{1,ij} +\beta_2 X_{2,ij} +\ldots+\beta_p X_{p,ij} + \zeta_j + \varepsilon_{ij} \\ &= (\beta_0+\zeta_j)+\beta_1 X_{1,ij} +\beta_2 X_{2,ij} +\ldots+\beta_p X_{p,ij} + \varepsilon_{ij} \qquad (i=1,\ldots,n_j, j=1,\ldots,J) \end{aligned}$$

  The intercept, $(\beta_0+\zeta_j)$, is not a constant, but rather depends on the manager (level-2 unit). It 'changes' from manager to manager.
- The intercept is a *random coefficient* (because managers are randomly drawn) and not associated with any explanatory variable, and thus the model is called a *random intercept model*.
- **ANOVA RE is in fact a RIM but without any covariate.**
- The variance of the intercept, $\Psi = \text{Var}(\zeta_j)$, is all that is estimated with regard to this quantity ζ_j, just like $\theta = \text{Var}(\varepsilon_{ij})$ is all that is estimated with regard to the 'remaining' (actual) error term, ε_{ij}.
- This clearly makes these random parameters different from any one of the β's, which are themselves estimated. Since the latter are population-specific, they are called *fixed parameters*, unlike the ζ and ε that are called *random effects*. The terms $\beta_s X_s$ (s = 2, …, p), or the β's, are called *fixed effects*. (This reference is often used more loosely for the β's.)

4. PTLV – Random Intercept Models

- A statistical model is a set of equation(s) with some distributional assumptions.
- For RIM,

$$y_{ij} = (\beta_0 + \zeta_j) + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \ldots + \beta_p X_{p,ij} + \varepsilon_{ij} \qquad (i=1,\ldots,n_j, j=1,\ldots,J)$$

  the assumptions are

  1) the ζ's are independent of one another (across managers) but remain the same across employees with same manager (i.e., constant within manager), since they are manager-specific;
  2) the ε's are worker-specific and vary across workers *and* managers (unlike the ζ's), being independent of one another across workers and managers;
  3) the ζ's and ε's are independent of each other (any ζ of any ε); and
  4) $E(\zeta_j \mid \underline{X}_{ij}) = E(\varepsilon_{ij} \mid \zeta_j, \underline{X}_{ij}) = 0$ (and hence also $E(\varepsilon_{ij} \mid \underline{X}_{ij}) = 0$); these two equations are referred to as *exogeneity assumptions*.

- The manger-specific error component ζ_j represents all effects of omitted manager characteristics upon the response *y*, or *unobserved heterogeneity*. Since ζ_j is the same for all subordinates to a given manager (j^{th}), ζ_j induces within-manager dependence among the total residuals e_{ij}. This is a new phenomenon for standard regression, which it does not deal with.
- with assumptions (i) through (iv),

$$E(y_{ij} \mid \zeta_j, \underline{X}_{ij}) = \beta_0 + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \ldots + \beta_p X_{p,ij} + \zeta_j .$$

- This above equation represents the *cluster-specific (unit-specific)* or *conditional regression* – given ζ and all the covariates (denoted for simplicity $\underline{X}$) – while averaged only over ε.
- Alternatively, the *population-averaged* or *marginal regression* is (averaged over ζ and ε, given the covariates $\underline{X}$):

$$E(y_{ij} \mid \underline{X}_{ij}) = \beta_0 + \beta_1 X_{1,ij} + \beta_2 X_{2,ij} + \ldots + \beta_p X_{p,ij} .$$

4. PTLV – Random Intercept Models

- Example

```
xtmixed masttest male mage hsgrad somecoll collgrad married black ||manager:

Performing EM optimization:
Performing gradient-based optimization:
Iteration 0:   log likelihood =  -34610.86
Iteration 1:   log likelihood = -34610.857
Computing standard errors:
Mixed-effects ML regression                     Number of obs      =      8604
Group variable: manager                         Number of groups   =      3978
                                                Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3
                                                Wald chi2(7)       =    494.88
Log likelihood = -34610.857                     Prob > chi2        =    0.0000
------------------------------------------------------------------------------
    masttest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   3.462295   .2735761    12.66   0.000     2.926095    3.998494
        mage |   .2144932   .0388718     5.52   0.000     .1383059    .2906804
      hsgrad |   2.952742   .7180955     4.11   0.000     1.545301    4.360184
    somecoll |   4.050778   .7775768     5.21   0.000     2.526756    5.574801
    collgrad |   4.900397   .7875138     6.22   0.000     3.356899    6.443896
     married |   2.917374   .7255747     4.02   0.000     1.495274    4.339474
       black |  -5.573554   .8170046    -6.82   0.000    -7.174854   -3.972255
       _cons |   86.30132   1.132518    76.20   0.000     84.08163    88.52102
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
manager: Identity            |
                   sd(_cons) |   9.945966    .179843      9.599654    10.30477
-----------------------------+------------------------------------------------
                sd(Residual) |    10.5647    .110056      10.35119    10.78263
------------------------------------------------------------------------------
LR test vs. linear regression: chibar2(01) =  1199.15 Prob >= chibar2 = 0.0000
```

4. PTLV – Random Intercept Models

- Comparing RIM

  $$y_{ij}=(\beta_0+\zeta_j)+\beta_1 X_{1,ij}+\beta_2 X_{2,ij}+\ldots+\beta_p X_{p,ij}+\varepsilon_{ij}$$

  with fully unconditional, *combined (mixed) model* associated with one-way ANOVA RE, i.e.

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

  which one is *better*?
- This question is answered using the proportional reduction in error variance (PREV) quantity that is defined as follows

  PREV = [EV(null model) – EV(full model)] / EV(null model)

  where EV(.) denotes 'error variance', i.e. EV = $\Psi + \theta$.
- The **full model** for the PREV is the currently considered RIM, while the **null model** is the ANOVA RE model (fully unconditional model).
  - Example

```
xtreg masttest, i(manager) mle
Iteration 0:   log likelihood = -34850.322   Iteration 1:   log likelihood = -34850.281
Random-effects ML regression                    Number of obs      =      8604
Group variable: manager                         Number of groups   =      3978
Random effects u_i ~ Gaussian                   Obs per group: min =         2
                                                               avg =       2.2
                                                               max =         3
                                                Wald chi2(0)       =      0.00
Log likelihood  = -34850.281                    Prob > chi2        =         .
------------------------------------------------------------------------------
    masttest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   99.94417   .2031043   492.08   0.000     99.54609    100.3422
-------------+----------------------------------------------------------------
    /sigma_u |   10.47977   .1836636                      10.12591      10.846
    /sigma_e |   10.74643   .1117388                      10.52964    10.96768
         rho |   .4874391   .0114188                      .4650901    .5098276
------------------------------------------------------------------------------
Likelihood-ratio test of sigma_u=0: chibar2(01)= 1315.66 Prob>=chibar2 = 0.000
```

PREV = (225.39 – 210.52)/225.39 = .066

- EV(null model) = $10.48^2 + 10.75^2$ = 225.39, EV(full model) = EV(RIM) = $9.95^2 + 10.56^2$ = 210.52

  i.e. 6.6% variance reduction is achieved when going from the ANOVA RE to the RIM.

4. PTLV – Random Intercept Models

- Hypotheses testing, e.g. testing significance of sets of predictors/explanatory variables, using the Wald tests.
  - ❖ Example

```
testparm married hsgrad
 ( 1)  [masttest]hsgrad = 0
 ( 2)  [masttest]married = 0
           chi2(  2) =   44.07
         Prob > chi2 =    0.0000
```

  the effect upon worker's mastery test score of manager marital status and education cannot be dispensed with (i.e. the hypothesis that the regression coefficients for manager age and education being both 0, is rejected).

- Predicted means (on the outcome variable) for level-2 or level-1 units
  - ❖ Example – in the case of RIM for married non-black managers of female employees, who have various levels of education.
    1) First generating a corresponding educational level variable

```
gen educ = hsgrad*1+somecoll*2+collgrad*3
label define ed 0 "no HS grad" 1 "HS grad" 2 "some Coll" 3 "Coll grad“
label values educ ed
```

    2) Then 'cross-tabulating' the predicted means by RIM

```
adjust male=0 married=1 black=0, by(educ) ci format(%4.0f)
--------------------------------------------------------------------------------------
     Dependent variable: masttest     Equation: masttest     Command: xtmixed
   Variables left as is: mage, hsgrad, somecoll, collgrad
Covariates set to value: male = 0, married = 1, black = 0
--------------------------------------------------------------------------------------
----------------------------------------------
      educ |         xb           lb           ub
----------+-----------------------------------
no HS grad |         94          [93           95]
   HS grad |         98          [97           99]
 some Coll |         99          [99          100]
 Coll grad |        101         [100          102]
----------------------------------------------
     Key:  xb         =  Linear Prediction
           [lb , ub]  =  [95% Confidence Interval]
```

4. PTLV – Model Adequacy Evaluation

- MLM does not have a 'saturated model' that could be used as a benchmark for evaluating overall model fit.
- Model adequacy is evaluated using residuals. In a 2-level modeling, there are 3 types of residuals
  1) Total residuals $\hat{e}_{ij} = y_{ij} - \hat{\beta}_o - \hat{\beta}_1 X_{1,ij} - \ldots - \hat{\beta}_p X_{p,ij}$ 2) Level-1 residuals $\hat{\varepsilon}_{ij} = \hat{e}_{ij} - \hat{\zeta}_j$
  3) Level-2 residuals $\hat{\zeta}_j$
- The level-2 residuals, ζ_j, can be obtained from *empirical Bayes estimates*. The Stata does this internally and automatically, using a main modeling module, 'gllamm'.
- Example – the same data set from the employee mastery test study.

```
gllamm masttest male mage hsgrad somecoll collgrad married black, i(manager) adapt
Running adaptive quadrature
Iteration 0:    log likelihood = -35207.583     Iteration 1:    log likelihood = -34924.634
Iteration 2:    log likelihood = -34624.349     Iteration 3:    log likelihood = -34610.913
Iteration 4:    log likelihood =  -34610.86     Iteration 5:    log likelihood =  -34610.86
Adaptive quadrature has converged, running Newton-Raphson
Iteration 0:   log likelihood =  -34610.86      Iteration 1:   log likelihood =  -34610.86
(backed up) Iteration 2:   log likelihood = -34610.857
number of level 1 units = 8604 number of level 2 units = 3978Condition Number = 186.0425
gllamm model
log likelihood = -34610.857
----------------------------------------------------------------------------
    masttest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+--------------------------------------------------------------
        male |   3.462295   .2735929    12.65   0.000     2.926063    3.998527
        mage |   .2144925   .0389559     5.51   0.000     .1381404    .2908447
      hsgrad |   2.952749   .7181249     4.11   0.000      1.54525    4.360248
    somecoll |   4.050788   .7776692     5.21   0.000     2.526585    5.574992
    collgrad |   4.900411    .787755     6.22   0.000      3.35644    6.444382
     married |   2.917371   .7255996     4.02   0.000     1.495222     4.33952
       black |  -5.573552    .817007    -6.82   0.000    -7.174856   -3.972248
       _cons |   86.30134   1.134018    76.10   0.000      84.0787    88.52397
------------------------------------------------------------------------------
Variance at level 1
------------------------------------------------------------------------------
   111.61281 (2.325413)
 Variances and covariances of random effects
------------------------------------------------------------------------------
 ***level 2 (manager)
    var(1): 98.922294 (3.5774213)
```

4. PTLV – Model Adequacy Evaluation-cont'd

- Example (cont'd)

  Store the associated estimates that will be handy for later use.

```
estimates store gllamm
```

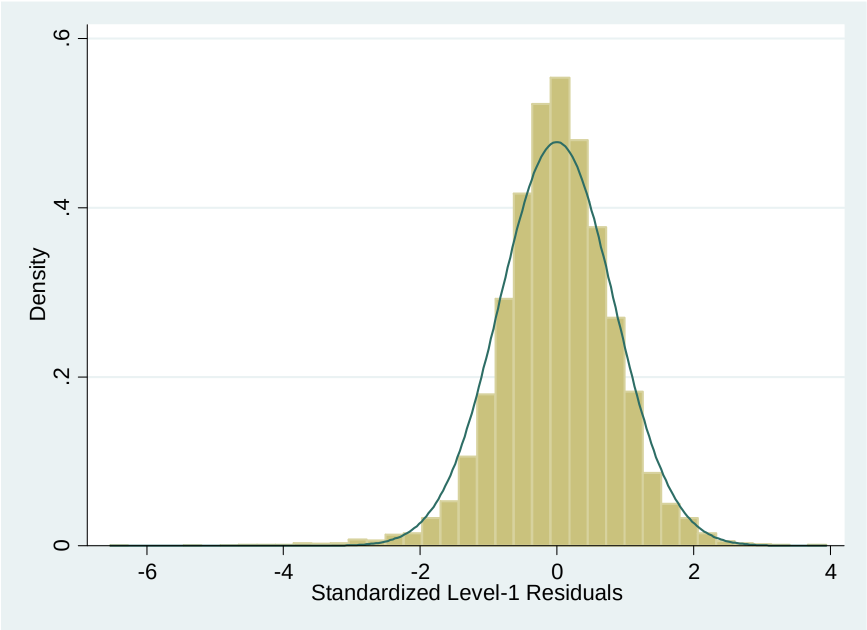
  Evaluate the fit of the model – looking at their histograms provided below.

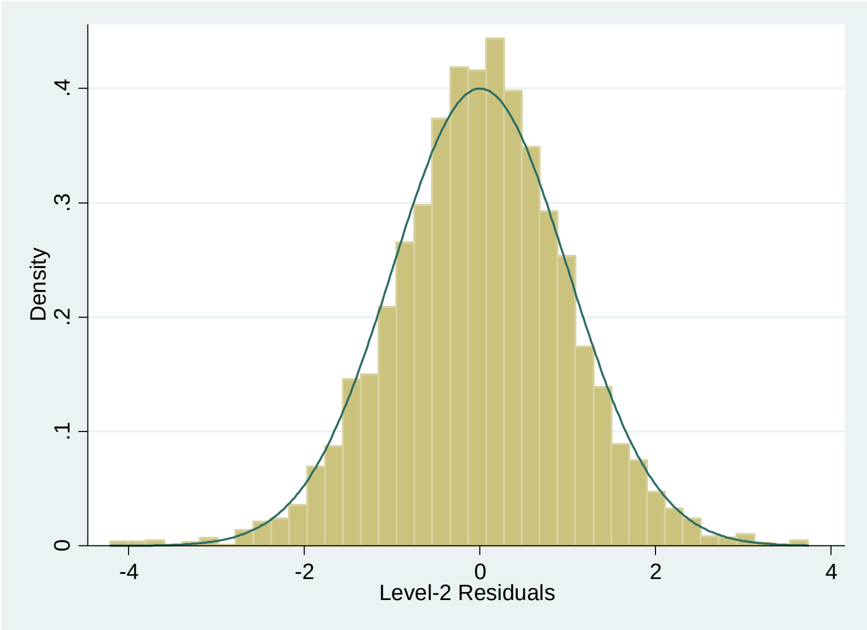
```
gllapred lev1, pearson
(residuals will be stored in lev1)
Non-adaptive log-likelihood: -34610.952
-3.461e+04 -3.461e+04
log-likelihood:-34610.857
histogram lev1, normal xtitle(Standardized Level-1 Residuals)
(bin=39, start=-6.5373143, width=.26869142)
gllapred lev2,ustd
histogram lev2m1, normal xtitle(Standardized Level-2 Residuals)
```

  These last commands produce the histograms of standardized level-1 and level-2 residuals

The slight deviation from normality in both plots could be possibly due to violations of some model assumptions. It would be recommendable to use a robust estimation procedure as below

```
gllamm, robust
```





4. PTLV – Model Adequacy Evaluation-cont'd

- Example (cont'd)

  Robust estimation procedure to overcome slight non-normality distribution in the residuals:

```
gllamm, robust
Non-adaptive log-likelihood: -34610.952
-3.461e+04 -3.461e+04
number of level 1 units = 8604          number of level 2 units = 3978
Condition Number = 186.0425

gllamm model

log likelihood = -34610.857
Robust standard errors
------------------------------------------------------------------------------
    masttest |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |   3.462295   .2771227    12.49   0.000     2.919145    4.005446
        mage |   .2144925   .0409445     5.24   0.000     .1342428    .2947422
      hsgrad |   2.952749   .7570516     3.90   0.000     1.468955    4.436543
    somecoll |   4.050788   .8204092     4.94   0.000     2.442816    5.658761
    collgrad |   4.900411   .8325151     5.89   0.000     3.268711    6.532111
     married |   2.917371   .7740334     3.77   0.000     1.400293    4.434448
       black |  -5.573552    .850151    -6.56   0.000    -7.239818   -3.907287
       _cons |   86.30134    1.19685    72.11   0.000     83.95555    88.64712
------------------------------------------------------------------------------
Variance at level 1
------------------------------------------------------------------------------
  111.61281 (3.7655232)
Variances and covariances of random effects
------------------------------------------------------------------------------
***level 2 (manager)
    var(1): 98.922294 (3.9585828)
------------------------------------------------------------------------------
```

  From this table, standard errors are larger than when the first fitted model with 'gllamm'. But they are 'safe' to use as far as finding out the individual predictor contribution is concerned to explain variability in the response variable.

5. RMAC

Robust Modeling Accounting for Clustering

5. RMAC

- RMAC, *robust modeling accounting for clustering* (Woodridge, 2009), is a modeling choice for the clustering effect when
  - dealing with hierarchical data
  - ICC is *not very pronounced, yet substantive*
  - the study design *needs* to take into account for the clustering
- A *Robust Modeling* is an alternative approach to standard statistical methods, such as those for estimating location, scale and regression parameters. It "produces estimators that are not unduly affected by small departures from the model assumptions under which these standard methods are usually derived: the standard methods are comparatively badly affected." (***Robust statistics*** – Wikipedia)

6. Summary

6. Summary

1. Statistical studies of data from social, behavioral, biomedical, business disciplines are dealing with studied subjects (units of analysis) that are clustered within higher-order units. MLM becomes an increasingly popular methodology of choice.
2. Application of classical, single-level statistical models, e.g. linear model, generalized linear model, "1st-generation" structural equation modeling, will in general produce
   1. mostly spuriously significant parameters
   2. short confidence intervals
   3. liberal statistical tests
3. Using aggregation and disaggregation to deal with nesting phenomenon doesn't deliver answers to the actual question of interest in MLM, if not misleading.
4. MLM offers a unified statistical methodology accounting for the lack of independence across level-1 units of analysis, being achieved by using an extension of OLS regression, viz. generalized least squares.
5. MLM is based on a large-sample statistical theory, and thus best used with 'large samples'. This is specifically a requirement with regard to number of higher-order units. For example, for 2-level modeling, *the units in level-2 has to be large*, e.g. 30+. And it is *not essential that the number of level-1 units be large, as long as that there is at least some level-2 units having 2+ cases in them*.
6. "MLM does not have routinely avaiable 'saturated model'…", "and hence no similar overall goodness of fit index(-es)."

Listing of Stata Commands

- Data input/output
  - `use “c:\path\data.dta”, clear`
  - `infile var1 var2 using “c:\path\data.dat”`
  - `save “c:\path\newdata.dta”, replace`
- Summary
  - `d                      /* display – print to screen the input */`
  - `sum                    /* summary */`
  - `graph twoway (scatter y_var x_var, mlabel(id_var)) (lfit y_var, x_var) if id_var <=8`
  - `list in 1/50`
  - `mat list e(V)          /* print to screen a matrix */`
- Analyses
  - `reg y_var x_var`
  - `xtmixed y_var x_var || id_var:, variance`
  - `predict re, reffects`
  - `predict se, reses`
  - `xtreg y_var, i(id_var) mle`

    `e.g. xtmixed income || region: || state:`
  - `testparm var1 var2     /* testing significance of individual var */`
- other