Day 2: Estimating models with multi-level data

Introduction to Multilevel Models
EUI Short Course 22–27 May, 2011
Prof. Kenneth Benoit

May 23, 2011

Preview: variance decomposition in MLMs

- Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

- We can model the dependence within subjects j by splitting ξ_{ij} into two components ζ_j and ϵ_{ij}:

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij}$$

- ζ_j represent level-2 effects, also known as "random intercepts", with variance ψ:

$$\zeta_j \sim N(0, \psi)$$

- ϵ_{ij} are level-1 errors, with variance θ

$$\epsilon_{ij} \sim N(0, \theta)$$

Basic assumptions of the Classical Linear Regression Model

1. Specification:
   - Relationship between X and Y *in the population* is linear: $E(Y) = X\beta$
   - No extraneous variables in X
   - No omitted independent variables
   - Parameters (β) are *constant*
2. $E(\epsilon) = 0$
3. Error terms:
   - $Var(\epsilon) = \sigma^2$, or homoskedastic errors
   - $E(r_{\epsilon_i,\epsilon_j}) = 0$, or no auto-correlation

Basic Assumptions of the CLRM (continued)

4. X is non-stochastic, meaning observations on independent variables are fixed in repeated samples
   - implies no *measurement error* in X
   - implies no serial correlation where a lagged value of Y would be used an independent variable
   - no *simultaneity* or *endogenous* X variables
5. $N > k$, or number of observations is greater than number of independent variables (in matrix terms: $\text{rank}(X) = k$), and no exact linear relationships exist in X
6. Normally distributed errors: $\epsilon | X \sim N(0, \sigma^2)$. Technically however this is a *convenience* rather than a strict assumption

Ordinary Least Squares (OLS)

- Objective: minimize $\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$, where
  - $\hat{Y}_i = b_0 + b_1 X_i$
  - error $e_i = (Y_i - \hat{Y}_i)$

$$
\begin{aligned}
b &= \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})} \\
&= \frac{\sum X_i Y_i}{\sum X_i^2}
\end{aligned}
$$

- The intercept is: $b_0 = \bar{Y} - b_1 \bar{X}$

- Closely related to ANOVA (sums of squares decomposition)

The “hat” matrix

- The hat matrix H is defined as:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ X\hat{\beta} &= X(X'X)^{-1}X'y \\ \hat{y} &= Hy\end{aligned}$$

- $H = X(X'X)^{-1}X'$ is called the *hat-matrix*
- Other important quantities, such as $\hat{y}$, $\sum e_i^2$ (RSS) can be expressed as functions of H
- Corrections for heteroskedastic errors (“robust” standard errors) involve manipulating H

Some important OLS properties to understand

Applies to $y = \alpha + \beta x + \epsilon$

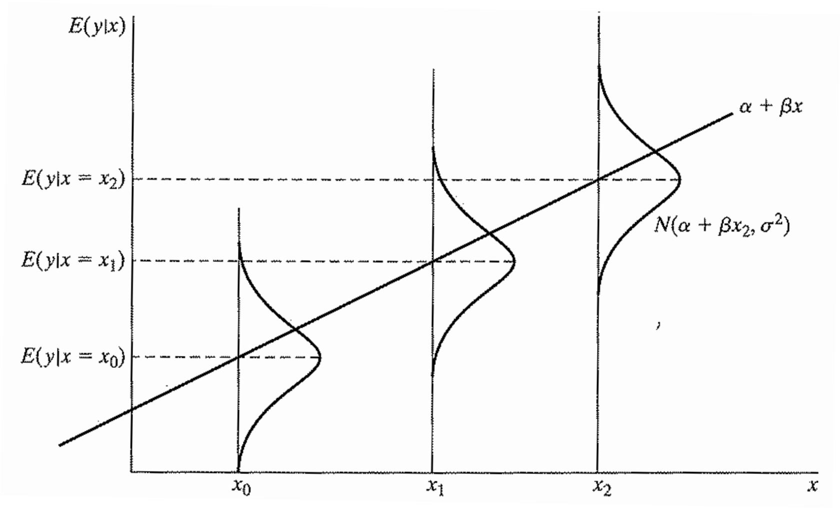
- If $\beta = 0$ and the only regressor is the intercept, then this is the same as regressing y on a column of ones, and hence $\alpha = \bar{y}$ — the mean of the observations
- If $\alpha = 0$ so that there is no intercept and one explanatory variable x, then $\beta = \frac{\sum xy}{\sum x^2}$
- If there is an intercept and one explanatory variable, then

$$\begin{aligned} \beta &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ &= \frac{\sum_i (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} \end{aligned}$$

Some important OLS properties (cont.)

- If the observations are expressed as deviations from their means, $y* = y - \bar{y}$ and $x^* = x - \bar{x}$, then $\beta = \sum x^* y^* / \sum x^{*2}$
- The intercept can be estimated as $\bar{y} - \beta\bar{x}$. This implies that the intercept is estimated by the value that causes the sum of the OLS residuals to equal zero.
- The mean of the $\hat{y}$ values equals the mean y values – together with previous properties, implies that the OLS regression line passes through the overall mean of the data points

Normally distributed errors



OLS in Stata

```
. use dail2002
(Ireland 2002 Dail Election - Candidate Spending Data)

. gen spendXinc = spend_total * incumb
(2 missing values generated)

. reg votes1st spend_total incumb minister spendXinc

      Source |       SS       df       MS              Number of obs =     462
-------------+------------------------------           F(  4,   457) =  229.05
       Model |  2.9549e+09     4   738728297           Prob > F      =  0.0000
    Residual |  1.4739e+09   457  3225201.58           R-squared     =  0.6672
-------------+------------------------------           Adj R-squared =  0.6643
       Total |  4.4288e+09   461  9607007.17           Root MSE      =  1795.9

------------------------------------------------------------------------------
    votes1st |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 spend_total |   .2033637   .0114807    17.71   0.000     .1808021    .2259252
      incumb |   5150.758   536.3686     9.60   0.000     4096.704    6204.813
    minister |   1260.001   474.9661     2.65   0.008      326.613     2193.39
   spendXinc |  -.1490399   .0274584    -5.43   0.000    -.2030003   -.0950794
       _cons |   469.3744   161.5464     2.91   0.004     151.9086    786.8402
------------------------------------------------------------------------------
```

Examples using the HSB data

The HSB dataset was originally compiled by Raudenbush and Bryk, and contains data on 7,185 students from 160 different schools.

Student-level variables:

mathach	student's mathematical ability (continuous)
ses	student's socioeconomic status (continuous)
min	binary variable for minority
female	binary variable for female

School level variables:

shoolid	numeric school ID
size	number of students attending that school
sector	whether school is public sector (0) or private (1)
disclim	disciplinary climate of the school (continuous)

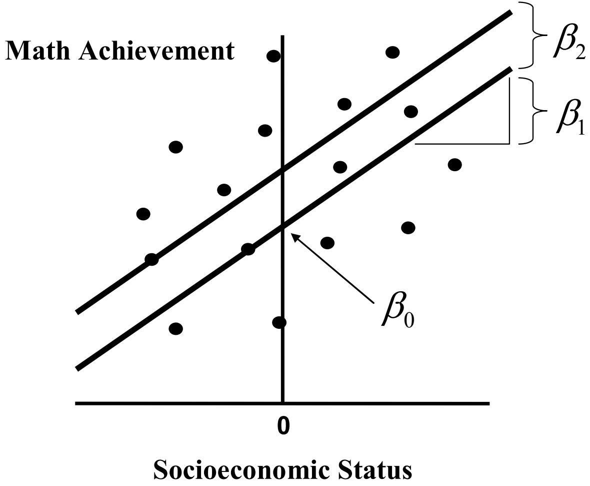
HSB data OLS model with two predictors

- Two predictors for math achievement will be socioeconomic status and whether the school is private or public
- Formula:

$$\texttt{mathach}_i = \beta_0 + \beta_1 \texttt{SES}_i + \beta_2 \texttt{sector}_i + \epsilon_i$$

- β_1 represents the average marginal effect of socioeconomic status on math achievement, holding sector constant
- β_2 represents the effect of of school sector holding socioeconomic status constant: the expected difference in achievement between public and private sector students who have the same socioeconomic status
- In this model only the distance of the effect can differ, as operationalized through the intercept
  - When $\texttt{sector}$=0, then $\texttt{mathach}_i = \beta_0 + \beta_1 \texttt{SES}_i + \epsilon_i$
  - When $\texttt{sector}$=0, then $\texttt{mathach}_i = (\beta_0 + \beta_2) + \beta_1 \texttt{SES}_i + \epsilon_i$

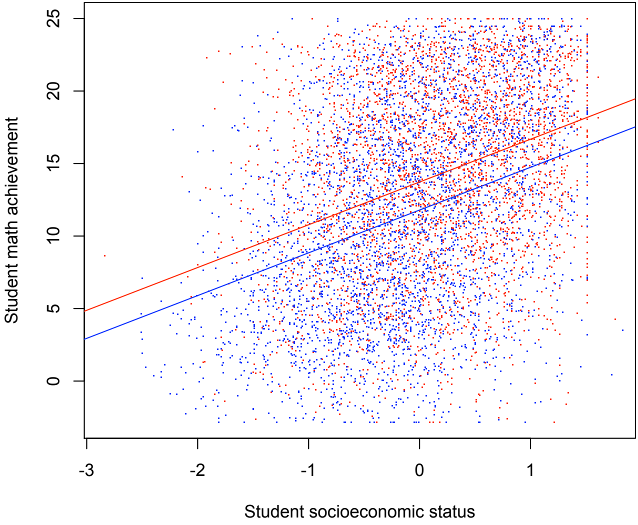
HSB data OLS model with two predictors cont.



HSB data regression: different intercepts

```
. use http://www.stata-press.com/data/mlmus2/hsb.dta, clear

. reg mathach ses sector

      Source |       SS       df       MS              Number of obs =    7185
-------------+------------------------------           F(  2,  7182) =  629.83
       Model |  50715.9161     2  25357.958           Prob > F      =  0.0000
    Residual |  289161.018  7182  40.2619073           R-squared     =  0.1492
-------------+------------------------------           Adj R-squared =  0.1490
       Total |  339876.934  7184  47.3102637           Root MSE      =  6.3452

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.948558   .0978306    30.14   0.000     2.756781    3.140334
      sector |   1.935013   .1524934    12.69   0.000     1.636081    2.233945
       _cons |   11.79325   .1061021   111.15   0.000     11.58526    12.00125
------------------------------------------------------------------------------
```

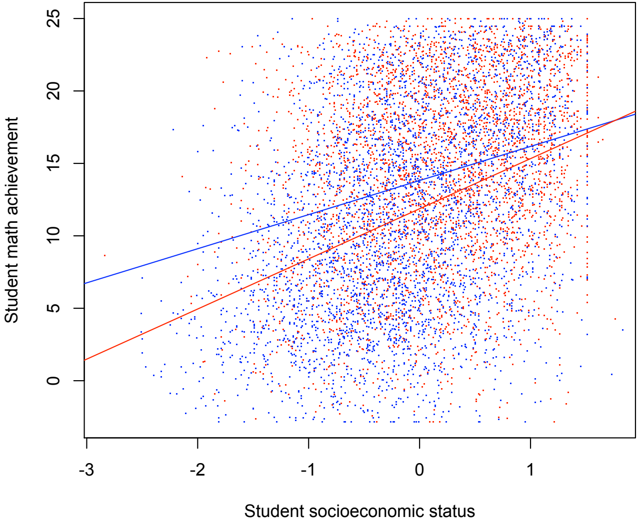
HSB data OLS model with “intercept dummy” for sector



HSB data regression: different slopes

```
. gen sesXsector = ses*sector

. reg mathach ses sector ses*sector

      Source |       SS       df       MS              Number of obs =    7185
-------------+------------------------------           F(  3,  7181) =  432.31
       Model |  51993.7695     3  17331.2565           Prob > F      =  0.0000
    Residual |  287883.165  7181  40.0895648           R-squared     =  0.1530
-------------+------------------------------           Adj R-squared =  0.1526
       Total |  339876.934  7184  47.3102637           Root MSE      =  6.3316

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   3.459632   .1331325    25.99   0.000     3.198653    3.720611
      sector |   1.949726    .152189    12.81   0.000     1.651391    2.248062
  sesXsector |  -1.105438   .1957985    -5.65   0.000     -1.48926   -.7216148
       _cons |   11.86764   .1066915   111.23   0.000      11.6585    12.07679
------------------------------------------------------------------------------
```

HSB data OLS model with different slopes



One way to “correct” multilevel problem: adjust SEs

- The key problem of having multilevel data is that our standard errors are wrong, since the assumption of conditional independence of the errors is violated because of the multilevel structure
- So one correction would be to “fix” the standard errors
  - The standard method is known as the White or Huber-White “modified sandwich estimator”, allowing for correction of clustered (heteroskedastic) errors
  - Stata has this built-in to most regression commands as an option: `, vec(cluster <clustervar>)`
- Works fine if we are only concerned with understanding the effects of the causal variables aggregated over all level 2 groups. It does not allow us to separate within- versus between-group effects, nor to examine how the effect of causal variables varies over level 2 groups

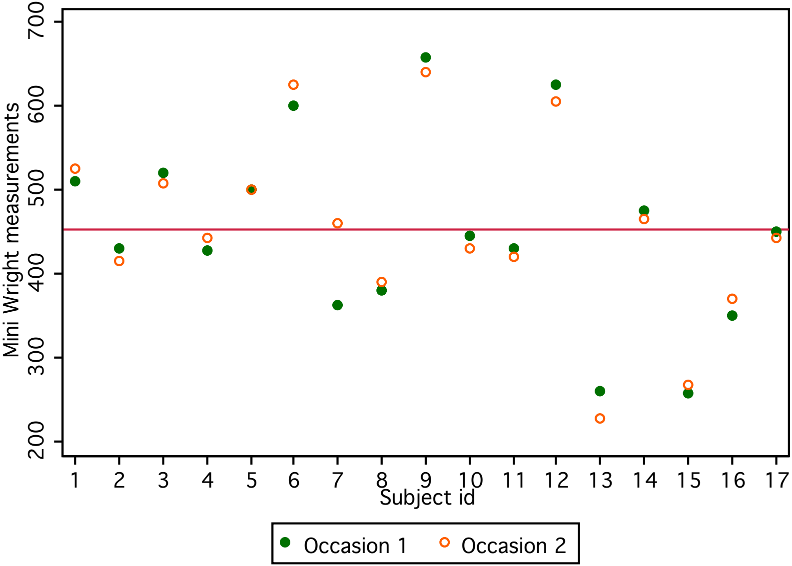
HSB data regression: different slopes

```
. reg mathach ses sector, vce(cluster schoolid)

Linear regression                                      Number of obs =    7185
                                                       F(  2,   159) =  354.53
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1492
                                                       Root MSE      =  6.3452

                             (Std. Err. adjusted for 160 clusters in schoolid)
------------------------------------------------------------------------------
             |               Robust
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.948558   .1279373    23.05   0.000     2.695882    3.201233
      sector |   1.935013   .3171766     6.10   0.000      1.30859    2.561436
       _cons |   11.79325   .2031455    58.05   0.000     11.39204    12.19447
------------------------------------------------------------------------------

. reg mathach ses sector

      Source |       SS       df       MS              Number of obs =    7185
-------------+------------------------------           F(  2,  7182) =  629.83
       Model |  50715.9161     2  25357.958            Prob > F      =  0.0000
    Residual |  289161.018  7182  40.2619073           R-squared     =  0.1492
-------------+------------------------------           Adj R-squared =  0.1490
       Total |  339876.934  7184  47.3102637           Root MSE      =  6.3452

------------------------------------------------------------------------------
     mathach |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |   2.948558   .0978306    30.14   0.000     2.756781    3.140334
      sector |   1.935013   .1524934    12.69   0.000     1.636081    2.233945
       _cons |   11.79325   .1061021   111.15   0.000     11.58526    12.00125
------------------------------------------------------------------------------
```

A multilevel dataset with two levels: Peak flow example

```
. list id wm1 wm2, clean

        id   wm1   wm2
  1.     1   512   525
  2.     2   430   415
  3.     3   520   508
  4.     4   428   444
  5.     5   500   500
  6.     6   600   625
  7.     7   364   460
  8.     8   380   390
  9.     9   658   642
 10.    10   445   432
 11.    11   432   420
 12.    12   626   605
 13.    13   260   227
 14.    14   477   467
 15.    15   259   268
 16.    16   350   370
 17.    17   451   443
```

A multilevel dataset with two levels: Peak flow example



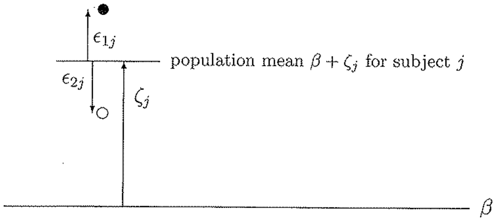
Variance components model

- Standard model without covariates:

$$y_{ij} = \beta + \xi_{ij}$$

- We can model the dependence within subjects j by splitting ξ_{ij} into two components ζ_j and ϵ_{ij}:

$$y_{ij} = \beta + \zeta_j + \epsilon_{ij}$$



Variance components model cont.

- In classic psychometric theory, $\beta + \zeta_j$ is subject j's "true score"
- The ζ_j are also called "random intercepts"
- We assume that the ζ_j are normally distributed:

$$\zeta_j \sim N(0, \psi)$$

  and that the ϵ_{ij} are normally distributed:

$$\epsilon_{ij} \sim N(0, \theta)$$

- The random intercept ζ_j can be thought of as a level-2 residual, with level-2 (between subject) variance ψ
- The random term ϵ_{ij} can be thought of as a level-1 residual (within subject) variance θ
- This also means that the observed responses y_{ij} are *conditionally independent* given ζ_j: $\text{Cor}(y_{ij}, y_{i'j} | \zeta_j) = 0$

Random intercept variance illustration

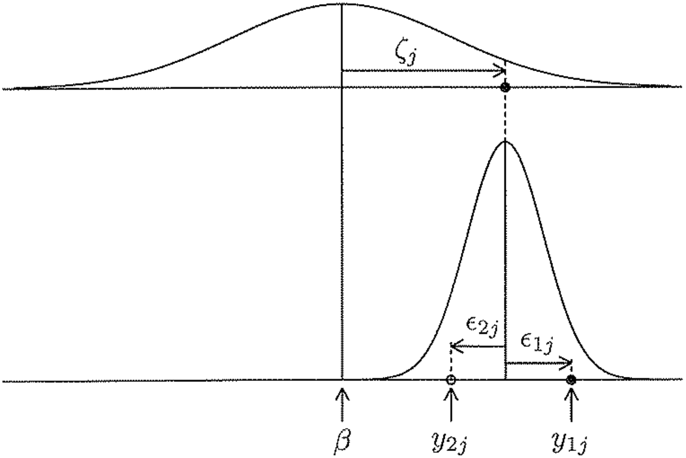


Figure 2.4: Illustration of distributions of error components for a subject j

Reliability

- Each overall error consists of the two error components ζ_j and ϵ_{ij}:

$$\xi_{ij} \equiv \zeta_j + \epsilon_{ij}$$

- The error components are independent, so it can be shown that the total variance is the sum of the between-subject and within-subject variances:

$$\begin{aligned} \mathrm{Var}(y_{ij}) &= \mathrm{Var}(\beta + \zeta_j + \epsilon_{ij}) \\ &= \mathrm{Var}(\beta) + \mathrm{Var}(\zeta_j + \epsilon_{ij}) \\ &= (0) + \psi + \theta \end{aligned}$$

- We can express the proportion of the total variance that is between subjects as:

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

- ρ can also be thought of as reliability of measurements for the same subjects j. It is also analogous to R^2 in that it represents the proportion of the total variance that is "explained" by subjects

Intraclass correlation

- ρ can also be interpreted as the marginal correlation between measurements on two occasions i and i' for the same subject
- So ρ aso represents within-cluster correlation
- We estimate the ICC using parameter estimates for ψ and θ:

$$\hat{\rho} = \frac{\hat{\psi}}{\hat{\psi} + \hat{\theta}}$$

- Can contrast the ICC with Pearson's r as:

$$r = \frac{\frac{1}{J-1}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{i\cdot})(y_{i'j} - \bar{y}_{i'\cdot})}{s_{y_i} s_{y_{i'}}}$$

- Pearson's r provides a measure of *relative agreement*, based on deviations of each i from their respective means
- ICC provides a measure of absolute agreement – and is therefore affected by linear transformations of the measurements