



## Bernoulli model for the THAI data

|                                      |   |
|--------------------------------------|---|
| 1. Introduction to HGLM models ..... | 1 |
| 2. Description of the model .....    | 2 |
| 3. Description of the data.....      | 3 |
| 4. Creating the command file.....    | 3 |
| 5. Interpreting the output .....     | 6 |

### 1. Introduction to HGLM models

Model specification for nonlinear analysis are specified via the **Basic Settings** dialog box as shown below. Six options are currently available. For the Binomial model and Poisson (variable exposure) the TRIAL or exposure variable is selected to the right of these two options; for the multinomial and ordinal models the number of categories should also be specified to the right of these two options.

If desired, and over-dispersion option is available for binomial and Poisson models. This option is not available with Laplace estimation. To specify over-dispersion, set the  $\sigma^2$  to **computed** on the **Estimation Settings** dialog box accessed via **Other Settings** on the main menu bar.

The nonlinear analysis is doubly iterative so the maximum number of macro iterations can be specified as well as the maximum number of micro iterations. Similarly, convergence criteria can be set for macro and micro iterations.

This is the second in a set of six examples illustrating HGLM models.

## 2. Description of the model

While the standard HLM uses a normal sampling model and an identity link function, the binary outcome model uses a binomial sampling model and a logit link. Only the level-1 models differ from the linear case.

Let  $Y_{ij}$  be the number of “successes” in  $m_{ij}$  trials. Then we write that

$$Y_{ij} | \phi_{ij} \sim B(m_{ij}, \phi_{ij}),$$

to denote that  $Y_{ij}$  has a binomial distribution with  $m_{ij}$  trials and probability of success  $\phi_{ij}$ . According to the binomial distribution, the expected value and variance of  $Y_{ij}$  are then

$$E(Y_{ij} | \phi_{ij}) = m_{ij}\phi_{ij} \quad \text{Var}(Y_{ij} | \phi_{ij}) = m_{ij}\phi_{ij}(1 - \phi_{ij}).$$

When  $m_{ij} = 1$ ,  $Y_{ij}$  may take on values of either zero or unity. This is a special case of the binomial distribution known as the Bernoulli distribution. HGLM allows estimation of models in which  $m_{ij} = 1$  (Bernoulli case) or  $m_{ij} > 1$  (other binomial cases).

For the Bernoulli case, the predicted value of the binary  $Y_{ij}$  is equal to the probability of a success,  $\phi_{ij}$ .

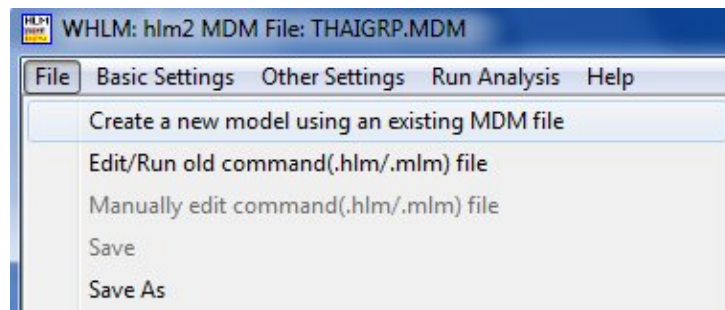
### 3. Description of the data

Data are from a national survey of primary education in Thailand (see Raudenbush & Bhumirat, 1992, for details), conducted in 1988, and yielding, for our analysis, complete data on 7516 sixth graders nested within 356 primary schools. Of interest is the probability that a child will repeat a grade during the primary years (REP1 = 1 if yes, 0 if no). It is hypothesized that the sex of the child (MALE = 1 if male, 0 if female), the child's pre-primary experience (PPED = 1 if yes, 0 if no), and the school mean SES (MSESC) will be associated with the probability of repetition. Every level-1 record corresponds to a student, with a single binary outcome per student, so the model type is Bernoulli. These data (level-1 and level-2) data files are UTHAIL1.SAV and THAI2.SAV.

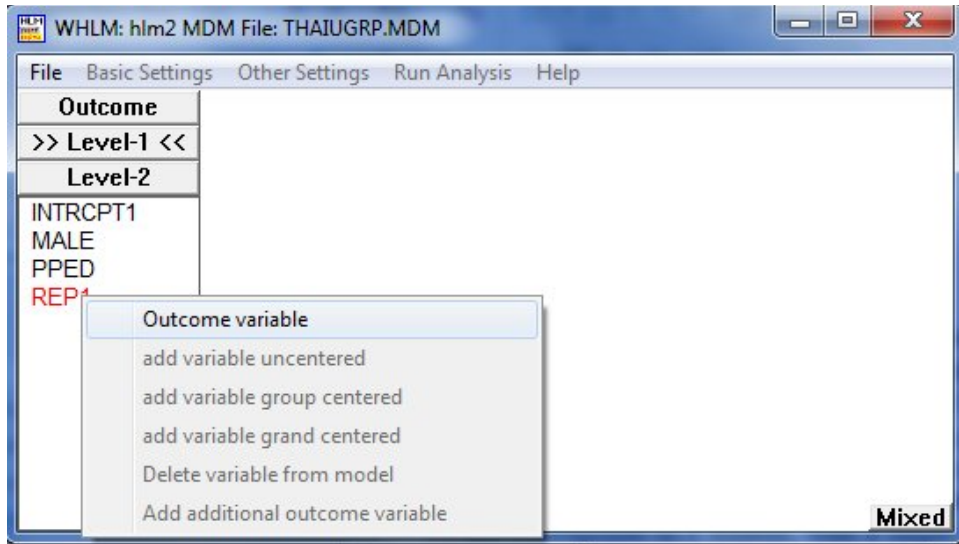
### 4. Creating the command file

The MDM file for a HGLM model is constructed in exactly the same way as for a linear model. The procedure is described in detail for the MDM and MDM data in other examples. Using the MDM file **THAIUGRP.MDM**, we set up the model as shown below.

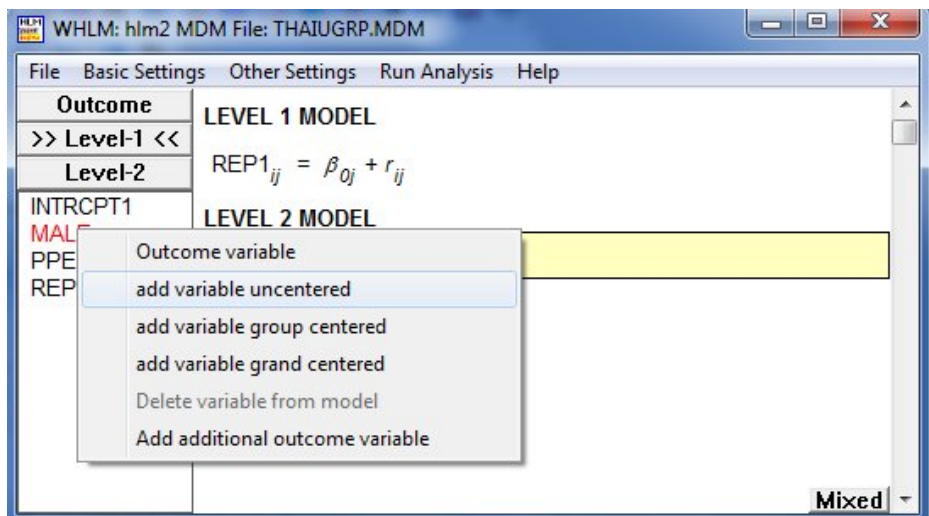
Start by selecting the **Create a new model using an existing MDM file** option from the **File** menu and open the MDM file **THAIGRP.MDM**.



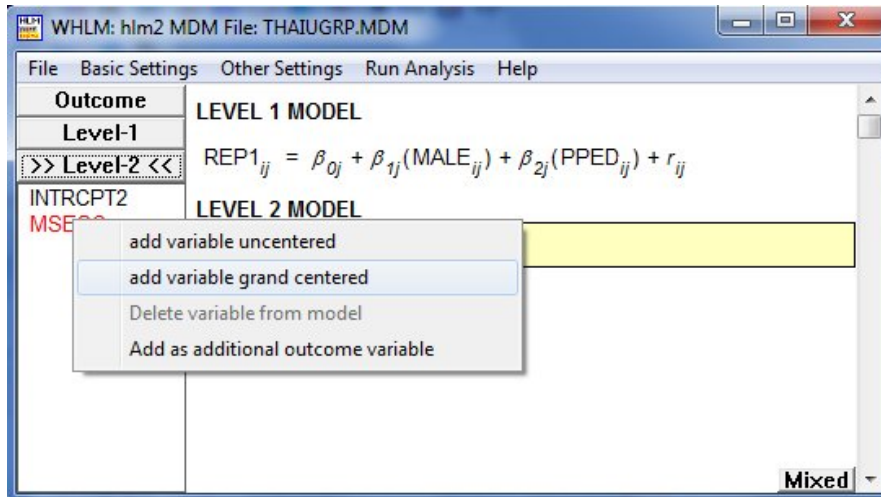
Select the outcome variable REP1 by clicking on the variable name at left and selecting **Outcome variable** from the pop-up menu.



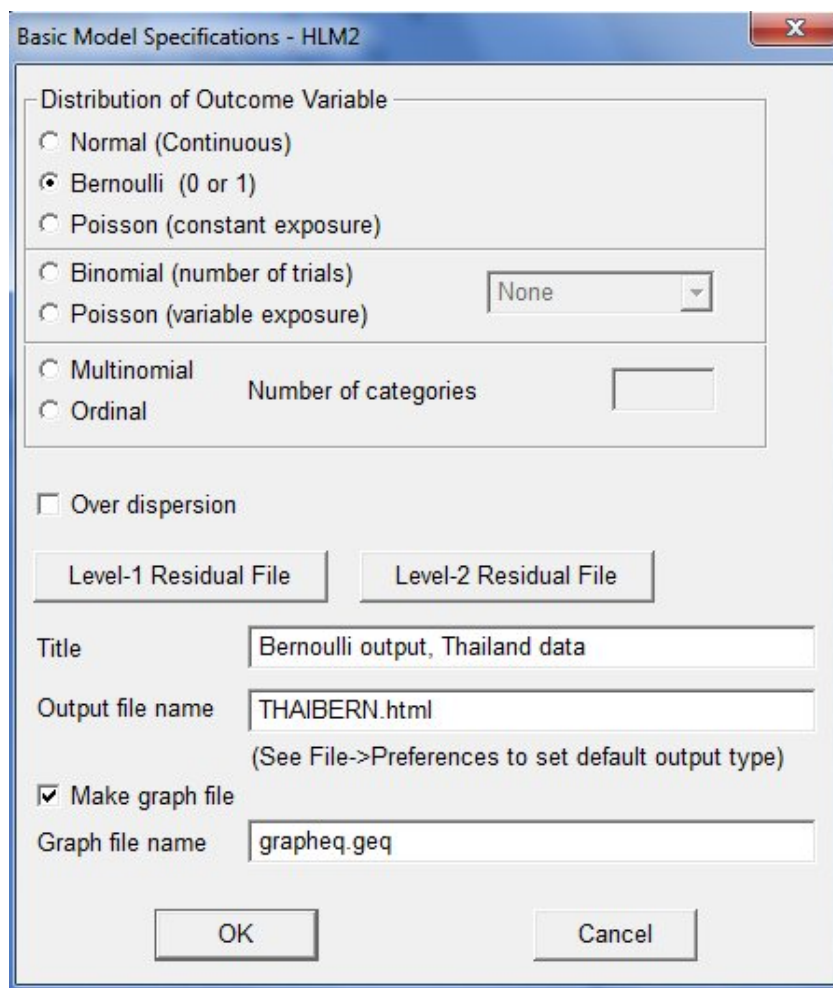
Next, add the variables MALE and PPED to the model by selecting the **add variable uncentered** option from the pop-up menu.



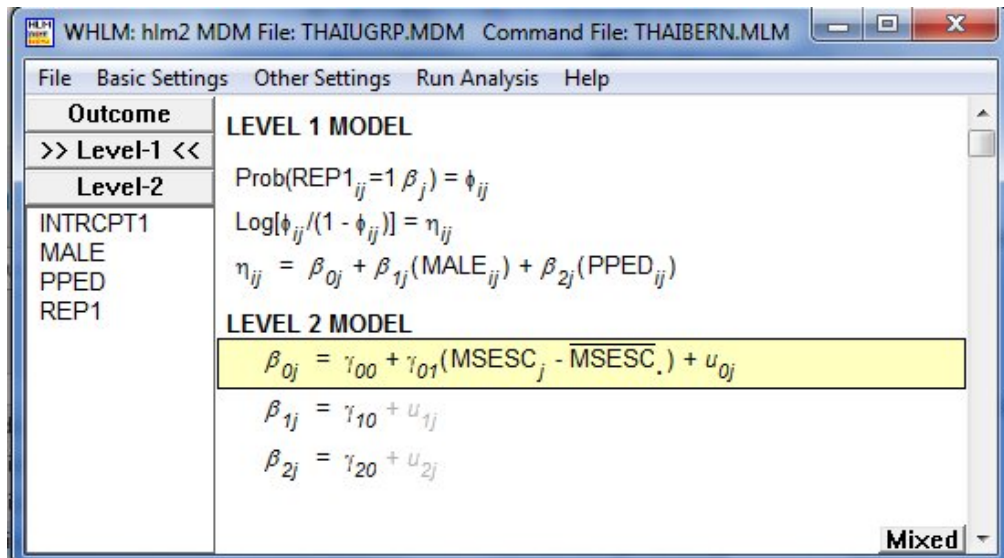
As a final step, include the level-2 predictor MDESC as **grand-mean centered** predictor on the level-2 intercept equation.



Click on **Basic Settings** on the main menu bar to indicate that the outcome has a bernoulli distribution.



Click **OK** to return to the main window. The model specification is now that of a bernoulli model using a logit link function.



Click **OK** to return to the main window and remember to save the command file prior to running the analysis.

## 5. Interpreting the output

### SPECIFICATIONS FOR THIS NONLINEAR HLM RUN

Problem Title: Bernoulli output, Thailand data

The data source for this run = THAIUGRP.MDM  
 The command file for this run = THAIBERN.HLM  
 Output file name = THAIBERN.HTML  
 The maximum number of level-1 units = 7516  
 The maximum number of level-2 units = 356  
 The maximum number of micro iterations = 20  
 Method of estimation: restricted PQL  
 Maximum number of macro iterations = 25

Distribution at Level-1: Bernoulli

The outcome variable is REP1

### Summary of the model specified

#### Level-1 Model

$$Prob(REP1_{ij}=1|\beta_j) = \phi_{ij}$$

$$\log[\phi_{ij}/(1 - \phi_{ij})] = \eta_{ij}$$

$$\eta_{ij} = \beta_{0j} + \beta_{1j}*(MALE_{ij}) + \beta_{2j}*(PPED_{ij})$$

Thus, the level-1 structural model is

$$\eta_{ij} = \log \left[ \frac{\phi_{ij}}{1 - \phi_{ij}} \right] = \beta_{0j} + \beta_{1j} (MALE)_{ij} + \beta_{2j} (PPED)_{ij}$$

### Level-2 Model

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(MSESC_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

MSESC has been centered around the grand mean.

And the level-2 structural model is

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (MSESC)_{ij} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20}$$

$$\text{Level-1 variance} = 1/[\phi_{ij} (1 - \phi_{ij})]$$

In the metric of the linearized dependent variable, the level-1 variance is the reciprocal of the Bernoulli variance,  $\phi_{ij} (1 - \phi_{ij})$ .

### Mixed Model

$$\eta_{ij} = \gamma_{00} + \gamma_{01}*MSESC_j + \gamma_{10}*MALE_{ij} + \gamma_{20}*PPED_{ij} + u_{0j}$$

Three sets of output results appear below: those for the normal linear model with identity link function, those for the unit-specific model with logit link function, and those for the population-average model with logit link. Typically, only the latter 2 sets of results will be relevant for drawing conclusions. The linear model with identity link is estimated simply to obtain starting values for the estimation of the models with logit link.

### Final Results for Linear Model with the Identity Link Function

$$\sigma^2 = 0.12181$$

$\tau$

$$\text{INTRCPT1, } \beta_0 \quad 0.01897$$

| Random level-1 coefficient | Reliability estimate |
|----------------------------|----------------------|
| INTRCPT1, $\beta_0$        | 0.749                |

The value of the log-likelihood function at iteration 6 = -2.413825E+003

**Estimation of fixed effects: (linear model with identity link function)**

| Fixed Effect              | Coefficient | Standard error | t-ratio | Approx. d.f. | p-value |
|---------------------------|-------------|----------------|---------|--------------|---------|
| For INTRCPT1, $\beta_0$   |             |                |         |              |         |
| INTRCPT2, $\gamma_{00}$   | 0.153756    | 0.010812       | 14.221  | 354          | <0.001  |
| MSESC, $\gamma_{01}$      | -0.033414   | 0.022465       | -1.487  | 354          | 0.138   |
| For MALE slope, $\beta_1$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{10}$   | 0.054131    | 0.008330       | 6.498   | 7158         | <0.001  |
| For PPED slope, $\beta_2$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{20}$   | -0.064613   | 0.010926       | -5.914  | 7158         | <0.001  |

**Results for Non-linear Model with the Logit Link Function  
Unit-Specific Model, PQL Estimation - (macro iteration 8)**

T  
INTRCPT1, $\beta_0$  1.29571

| Random level-1 coefficient | Reliability estimate |
|----------------------------|----------------------|
| INTRCPT1, $\beta_0$        | 0.682                |

The value of the log-likelihood function at iteration 2 = -1.001031E+004

**Final estimation of fixed effects: (Unit-specific model)**

| Fixed Effect              | Coefficient | Standard error | t-ratio | Approx. d.f. | p-value |
|---------------------------|-------------|----------------|---------|--------------|---------|
| For INTRCPT1, $\beta_0$   |             |                |         |              |         |
| INTRCPT2, $\gamma_{00}$   | -2.046961   | 0.093985       | -21.780 | 354          | <0.001  |
| MSESC, $\gamma_{01}$      | -0.254412   | 0.193319       | -1.316  | 354          | 0.189   |
| For MALE slope, $\beta_1$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{10}$   | 0.508561    | 0.073935       | 6.879   | 7158         | <0.001  |
| For PPED slope, $\beta_2$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{20}$   | -0.594375   | 0.095962       | -6.194  | 7158         | <0.001  |

| Fixed Effect              | Coefficient | Odds Ratio | Confidence Interval |
|---------------------------|-------------|------------|---------------------|
| For INTRCPT1, $\beta_0$   |             |            |                     |
| INTRCPT2, $\gamma_{00}$   | -2.046961   | 0.129127   | (0.107,0.155)       |
| MSESC, $\gamma_{01}$      | -0.254412   | 0.775372   | (0.530,1.134)       |
| For MALE slope, $\beta_1$ |             |            |                     |
| INTRCPT2, $\gamma_{10}$   | 0.508561    | 1.662897   | (1.439,1.922)       |
| For PPED slope, $\beta_2$ |             |            |                     |
| INTRCPT2, $\gamma_{20}$   | -0.594375   | 0.551908   | (0.457,0.666)       |



**Final estimation of fixed effects  
(Unit-specific model with robust standard errors)**

| Fixed Effect              | Coefficient | Standard error | t-ratio | Approx. d.f. | p-value |
|---------------------------|-------------|----------------|---------|--------------|---------|
| For INTRCPT1, $\beta_0$   |             |                |         |              |         |
| INTRCPT2, $\gamma_{00}$   | -2.046961   | 0.094872       | -21.576 | 354          | <0.001  |
| MSESC, $\gamma_{01}$      | -0.254412   | 0.204048       | -1.247  | 354          | 0.213   |
| For MALE slope, $\beta_1$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{10}$   | 0.508561    | 0.075994       | 6.692   | 7158         | <0.001  |
| For PPED slope, $\beta_2$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{20}$   | -0.594375   | 0.094840       | -6.267  | 7158         | <0.001  |

| Fixed Effect              | Coefficient | Odds Ratio | Confidence Interval |
|---------------------------|-------------|------------|---------------------|
| For INTRCPT1, $\beta_0$   |             |            |                     |
| INTRCPT2, $\gamma_{00}$   | -2.046961   | 0.129127   | (0.107,0.156)       |
| MSESC, $\gamma_{01}$      | -0.254412   | 0.775372   | (0.519,1.158)       |
| For MALE slope, $\beta_1$ |             |            |                     |
| INTRCPT2, $\gamma_{10}$   | 0.508561    | 1.662897   | (1.433,1.930)       |
| For PPED slope, $\beta_2$ |             |            |                     |
| INTRCPT2, $\gamma_{20}$   | -0.594375   | 0.551908   | (0.458,0.665)       |

**Final estimation of variance components**

| Random Effect   | Standard Deviation | Variance Component | d.f. | $\chi^2$   | p-value |
|-----------------|--------------------|--------------------|------|------------|---------|
| INTRCPT1, $u_0$ | 1.13829            | 1.29571            | 354  | 1431.43082 | <0.001  |

**Results for Population-Average Model**

The value of the log-likelihood function at iteration 2 = -1.010987E+004

**Final estimation of fixed effects: (Population-average model)**

| Fixed Effect              | Coefficient | Standard error | t-ratio | Approx. d.f. | p-value |
|---------------------------|-------------|----------------|---------|--------------|---------|
| For INTRCPT1, $\beta_0$   |             |                |         |              |         |
| INTRCPT2, $\gamma_{00}$   | -1.748402   | 0.087969       | -19.875 | 354          | <0.001  |
| MSESC, $\gamma_{01}$      | -0.283620   | 0.185179       | -1.532  | 354          | 0.127   |
| For MALE slope, $\beta_1$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{10}$   | 0.446546    | 0.066993       | 6.666   | 7158         | <0.001  |
| For PPED slope, $\beta_2$ |             |                |         |              |         |
| INTRCPT2, $\gamma_{20}$   | -0.536378   | 0.088479       | -6.062  | 7158         | <0.001  |

| Fixed Effect              | Coefficient | Odds Ratio | Confidence Interval |
|---------------------------|-------------|------------|---------------------|
| For INTRCPT1, $\beta_0$   |             |            |                     |
| INTRCPT2, $\gamma_{00}$   | -1.748402   | 0.174052   | (0.146,0.207)       |
| MSESC, $\gamma_{01}$      | -0.283620   | 0.753053   | (0.523,1.084)       |
| For MALE slope, $\beta_1$ |             |            |                     |
| INTRCPT2, $\gamma_{10}$   | 0.446546    | 1.562905   | (1.371,1.782)       |
| For PPED slope, $\beta_2$ |             |            |                     |
| INTRCPT2, $\gamma_{20}$   | -0.536378   | 0.584863   | (0.492,0.696)       |

Notice that the results for the population-average model are quite similar to the results for the unit-specific model except in the case of the intercept. The intercept in the population-average model in this case is the expected log-odds of repetition for a person with values of zero on the predictors (and therefore, for a female without pre-primary experience attending a school of average SES). In this case, this expected log-odds corresponds to a probability of  $1/(1 + \exp\{1.748402\}) = .148$ , which is the “population-average” repetition rate for this group. In contrast, the unit-specific intercept is the expected log-odds of repetition rate for the same kind of student, but one who attends a school that not only has a mean SES of 0, but also has a random effect of zero (that is, a school with a “typical” repetition rate for the school of its type). This conditional expected log-odds is  $-2.046961$ , corresponding to a probability of  $1/(1 + \exp\{2.046961\}) = .114$ . Thus the probability of repetition is lower in a school with a random effect of zero than the average in the population of schools having mean SES of zero taken as a whole. This is a typical result. Population-average probabilities will be closer to .50 (than will the corresponding unit-specific probabilities).

One final set of results is printed out: population-average results with robust standard errors (below). Note that the robust standard errors in this case are very similar to the model-based standard errors, with a slight increase for the level-2 predictor and slight decreases for level-1 predictors. Results for other data may not follow this pattern.

**Final estimation of fixed effects**  
**(Population-average model with robust standard errors)**

| Fixed Effect                                | Coefficient | Standard error | t-ratio | Approx. d.f. | p-value |
|---|-------------|----------------|---------|--------------|---------|
| <b>For INTRCPT1, <math>\beta_0</math></b>   |             |                |         |              |         |
| INTRCPT2, $\gamma_{00}$                     | -1.748402   | 0.082158       | -21.281 | 354          | <0.001  |
| MSESC, $\gamma_{01}$                        | -0.283620   | 0.196005       | -1.447  | 354          | 0.149   |
| <b>For MALE slope, <math>\beta_1</math></b> |             |                |         |              |         |
| INTRCPT2, $\gamma_{10}$                     | 0.446546    | 0.062788       | 7.112   | 7158         | <0.001  |
| <b>For PPED slope, <math>\beta_2</math></b> |             |                |         |              |         |
| INTRCPT2, $\gamma_{20}$                     | -0.536378   | 0.082221       | -6.524  | 7158         | <0.001  |

| Fixed Effect                                | Coefficient | Odds Ratio | Confidence Interval |
|---|-------------|------------|---------------------|
| <b>For INTRCPT1, <math>\beta_0</math></b>   |             |            |                     |
| INTRCPT2, $\gamma_{00}$                     | -1.748402   | 0.174052   | (0.148,0.205)       |
| MSESC, $\gamma_{01}$                        | -0.283620   | 0.753053   | (0.512,1.107)       |
| <b>For MALE slope, <math>\beta_1</math></b> |             |            |                     |
| INTRCPT2, $\gamma_{10}$                     | 0.446546    | 1.562905   | (1.382,1.768)       |
| <b>For PPED slope, <math>\beta_2</math></b> |             |            |                     |
| INTRCPT2, $\gamma_{20}$                     | -0.536378   | 0.584863   | (0.498,0.687)       |