Poisson model (equal exposure) for the THAI data

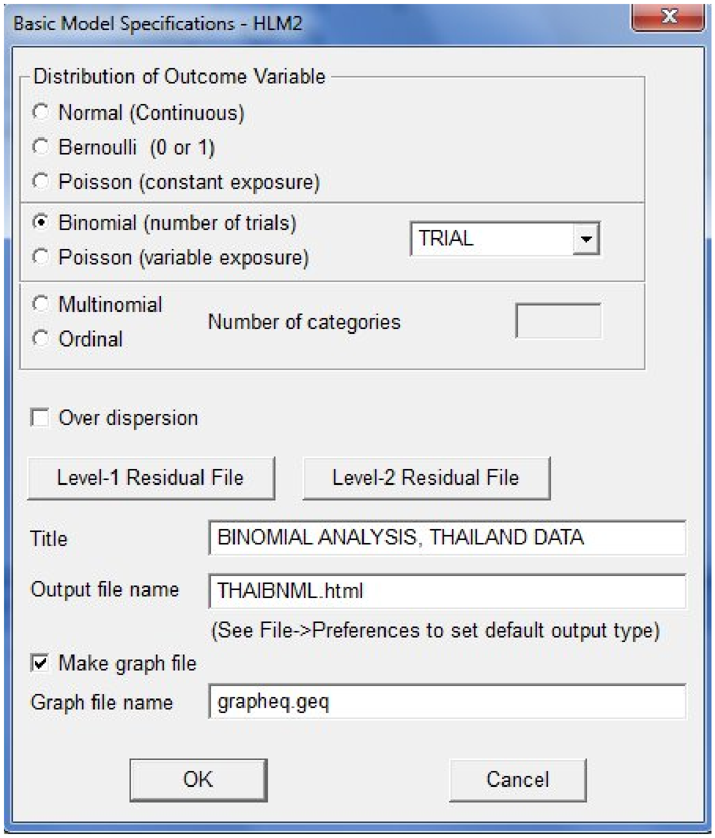
1. Introduction to HGLM models

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 σ^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 third in a set of six examples illustrating HGLM models.



2. Description of the model

For count data, we use a Poisson sampling model and a log link function. Let Y_{ij} be the number of events occurring during an interval of time having length m_{ij}. For example, Y_{ij} could be the number of crimes a person *i* from group *j* commits during five years, so that m_{ij} = 5. The time-interval of m_{ij} units may be termed the "exposure." Then we write that

$$Y_{ij} \mid \lambda_{ij} \sim P(m_{ij}, \lambda_{ij})$$

to denote that Y_{ij} has a Poisson distribution with exposure m_{ij} and event rate λ_j . According to the Poisson distribution, the expected value and variance of Y_{ij} are then

$$E(Y_{ij} \mid \lambda_{ij}) = m_{ij}\lambda_{ij} \qquad Var(Y_{ij} \mid \lambda_{ij}) = m_{ij}\lambda_{ij}.$$

The exposure m_{ij} need not be a measure of time. For example, if Y_{ij} is the number of bombs dropping on neighborhood *i* of city *j* during a war, m_{ij} could be the area of that neighborhood. A common case arises when, for each *i* and *j,* the exposure is the same (*e.g.*, Y_{ij} is the number of crimes committed during one year for each person *i* within each neighborhood *j*). In this case, we set m_{ij} = 1 for simplicity. HGLM allows estimation of models in which m_{ij} = 1 or m_{ij} ³ 1.

According to our level-1 model, the predicted value of Y_{ij} when m_{ij} = 1 will be the event rate λ_{ij}.

HGLM uses the log link function when the level-1 sampling model is Poisson, that is

$$\eta_{ij} = \log(\lambda_{ij}).$$

In words, η_{ij} is the log of the event rate. Thus, if the event rate, λ_{ij}, is one, the log is zero. When the event rate is less than one, the log is negative; when the event rate is greater than one, the log is positive. Thus, while λ_{ij} is constrained to be non-negative, η_{ij} can take on any real value. It is possible to generate a predicted log-event rate (η_{ij}) for any case. Such a predicted log-event rate can be converted to an event rate by computing $\lambda_{ij} = \text{event rate} = \exp(\eta_{ij})$. Clearly, whatever the value of η_{ij}, λ_{ij} will be non-negative.

3. Description of the data

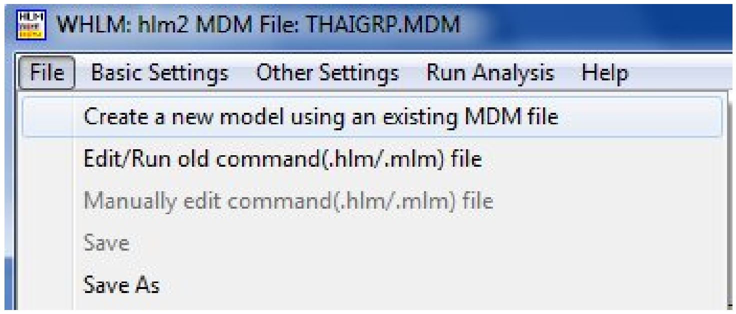
Data are from a national survey of primary education in Thailand (see Raudenbush & Bhumirat, 1992, for details), conducted in 1988, and yielding, for our analysis, complete data on 7516 sixth graders nested within 356 primary schools. Of interest is the probability that a child will repeat a grade during the primary years (REP1 = 1 if yes, 0 if no). It is hypothesized that the sex of the child (MALE = 1 if male, 0 of 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.

Suppose that the outcome variable is the number of days absent during the previous year rather than grade repetition. This outcome would be a non-negative integer, that is, a count rather than a dichotomy. Thus, the Poisson model with a log link would be a reasonable choice for the model. Notice that the time interval during which the absences could accumulate, that is, one year, would be the same for each student. We call this a case of "equal exposure," meaning that each level-1 case had an "equal opportunity" to accumulate absences.

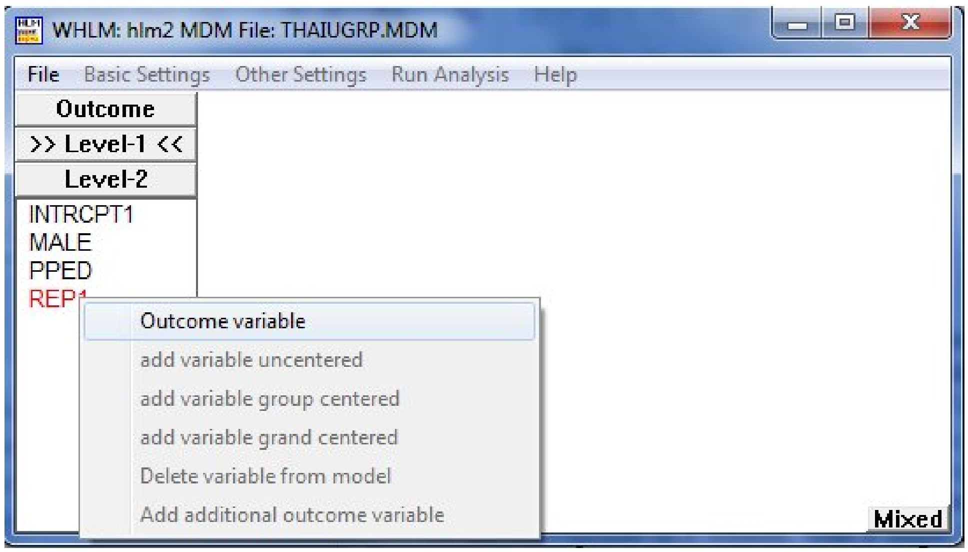
4. Creating the command file

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 **THAIGRP.MDM**, we set up the model as shown below.

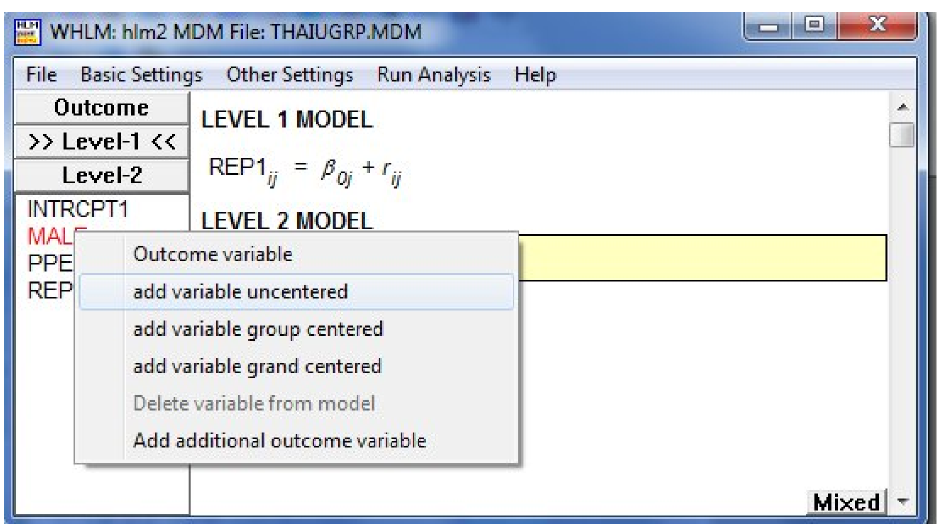
Start by selecting the **Create a new model using an existing MDM file** option from the **File** menu and open the MDM file **THAIGRP.MDM**.



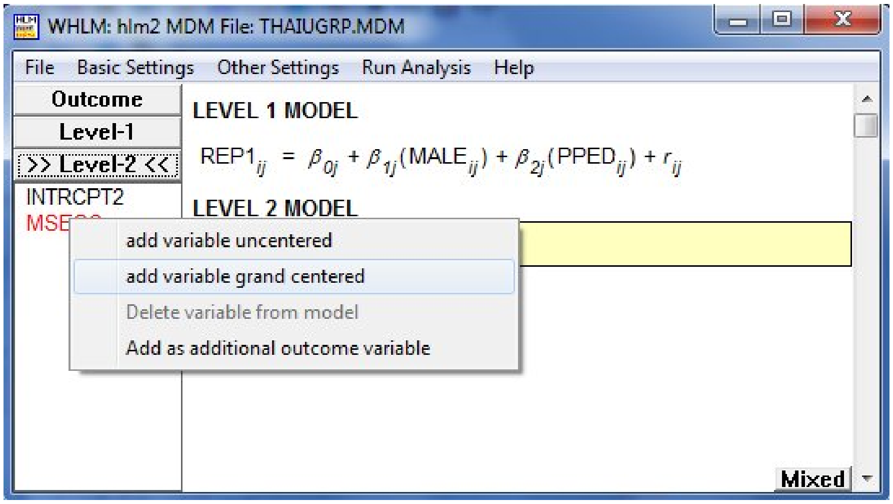
Select the outcome variable REP1 by clicking on the variable name at left and selecting **Outcome variable** from the pop-up menu.



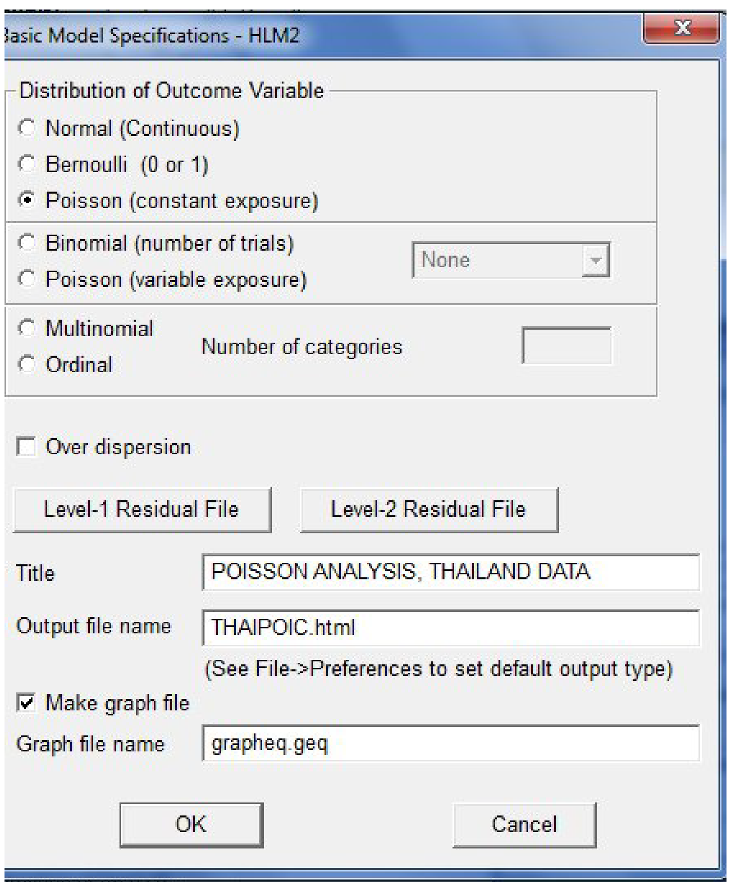
Next, add the variables MALE and PPED to the model by selecting the **add variable uncentered** option from the pop-up menu.



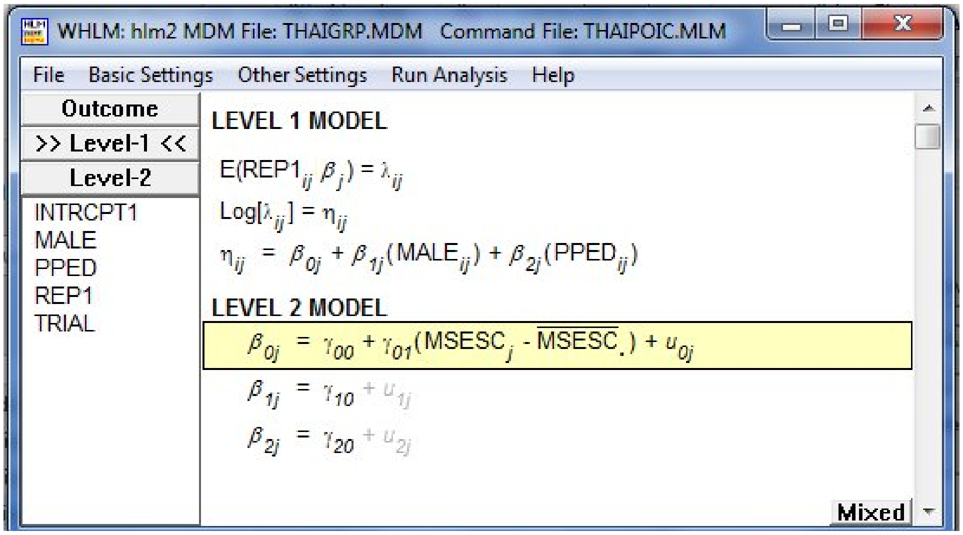
As a final step, include the level-2 predictor MSESC as **grand-mean centered** predictor on the level-2 intercept equation.



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



Click **OK** to return to the main window. The model specification is now that of a Poisson model.



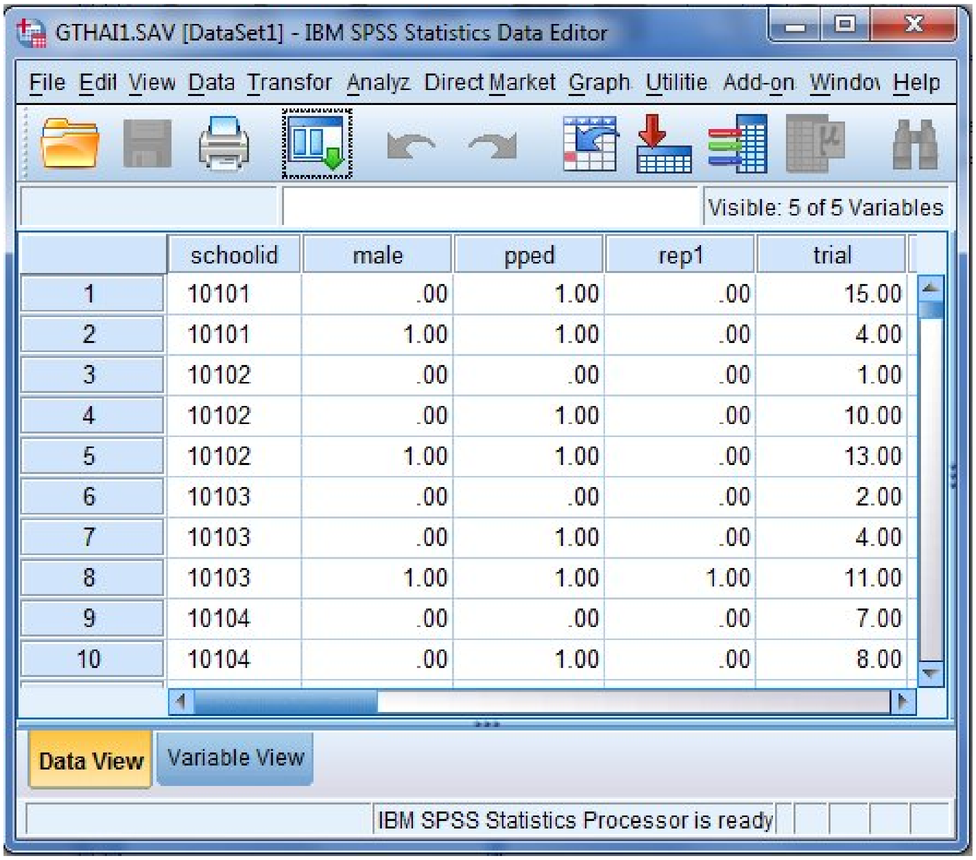
Click **OK** to return to the main window and remember to save the command file prior to running the analysis.

5. Interpreting the output

In the case of the Poisson model, the expected rate of repetition λ_{ij} can be expressed as

$$\hat{\lambda}_{ij} = \exp\left[\hat{\gamma}_{00} + \hat{\gamma}_{01}\left(MSESC_j - \overline{MSESC.}\right) + \hat{\gamma}_{10}\, MALE_{ij} + \hat{\gamma}_{20}\, PPED_{ij}\right]$$

Data for the first 10 observations are shown below.



	schoolid	male	pped	rep1	trial
1	10101	.00	1.00	.00	15.00
2	10101	1.00	1.00	.00	4.00
3	10102	.00	.00	.00	1.00
4	10102	.00	1.00	.00	10.00
5	10102	1.00	1.00	.00	13.00
6	10103	.00	.00	.00	2.00
7	10103	.00	1.00	.00	4.00
8	10103	1.00	1.00	1.00	11.00
9	10104	.00	.00	.00	7.00
10	10104	.00	1.00	.00	8.00

Output for this model again include unit-specific and population-average results.

τ
INTRCPT1,β_0 0.73034

Standard error of τ
INTRCPT1,β_0 0.08662

Approximate confidence intervals of tau variances
INTRCPT1 : (0.578,0.922)

Random level-1 coefficient	Reliability estimate
INTRCPT1,β_0	0.613

The value of the log-likelihood function at iteration 2 = -1.760082E+03

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

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-0.358667	0.074569	-4.810	354	<0.001
MSESC, γ_{01}	-0.404192	0.152612	-2.649	354	0.008
For MALE slope, β_1					
INTRCPT2, γ_{10}	0.377973	0.063185	5.982	739	<0.001
For PPED slope, β_2					
INTRCPT2, γ_{20}	-0.234582	0.070324	-3.336	739	<0.001

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	-0.358667	0.698607	(0.603,0.809)
MSESC, γ_{01}	-0.404192	0.667516	(0.494,0.901)
For MALE slope, β_1			
INTRCPT2, γ_{10}	0.377973	1.459324	(1.289,1.652)
For PPED slope, β_2			
INTRCPT2, γ_{20}	-0.234582	0.790901	(0.689,0.908)

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, β_0					
INTRCPT2, γ_{00}	-0.358667	0.081440	-4.404	354	<0.001
MSESC, γ_{01}	-0.404192	0.169526	-2.384	354	0.018
For MALE slope, β_1					
INTRCPT2, γ_{10}	0.377973	0.068249	5.538	739	<0.001
For PPED slope, β_2					
INTRCPT2, γ_{20}	-0.234582	0.101980	-2.300	739	0.022

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	-0.358667	0.698607	(0.595,0.820)
MSESC, γ_{01}	-0.404192	0.667516	(0.478,0.932)
For MALE slope, β_1			
INTRCPT2, γ_{10}	0.377973	1.459324	(1.276,1.669)
For PPED slope, β_2			
INTRCPT2, γ_{20}	-0.234582	0.790901	(0.647,0.966)

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	0.85460	0.73034	354	1362.92423	<0.001

Results for Population-Average Model

The value of the log-likelihood function at iteration 3 = -1.868414E+03

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

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-0.100642	0.069406	-1.450	354	0.148
MSESC, γ_{01}	-0.425816	0.145853	-2.919	354	0.004

For MALE slope, β_1					
INTRCPT2, γ_{10}	0.378377	0.058805	6.434	739	<0.001
For PPED slope, β_2					
INTRCPT2, γ_{20}	-0.248909	0.062770	-3.965	739	<0.001

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	-0.100642	0.904256	(0.789,1.037)
MSESC, γ_{01}	-0.425816	0.653237	(0.490,0.870)
For MALE slope, β_1			
INTRCPT2, γ_{10}	0.378377	1.459914	(1.301,1.639)
For PPED slope, β_2			
INTRCPT2, γ_{20}	-0.248909	0.779651	(0.689,0.882)

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

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	-0.100642	0.072801	-1.382	354	0.168
MSESC, γ_{01}	-0.425816	0.181892	-2.341	354	0.020
For MALE slope, β_1					
INTRCPT2, γ_{10}	0.378377	0.059110	6.401	739	<0.001
For PPED slope, β_2					
INTRCPT2, γ_{20}	-0.248909	0.082665	-3.011	739	0.003

Fixed Effect	Coefficient	Event Rate Ratio	Confidence Interval
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	-0.100642	0.904256	(0.784,1.043)
MSESC, γ_{01}	-0.425816	0.653237	(0.457,0.934)
For MALE slope, β_1			
INTRCPT2, γ_{10}	0.378377	1.459914	(1.300,1.640)
For PPED slope, β_2			
INTRCPT2, γ_{20}	-0.248909	0.779651	(0.663,0.917)

Substituting the estimates obtained from the output file we can express the expected rate of repetition as

$$\hat{\lambda}_{ij} = \exp\left[-0.100642 - 0.425816\left(MSESC_j - \overline{MSESC.}\right) + 0.378377 MALE_{ij} - 0.248909 PPED_{ij}\right]$$