

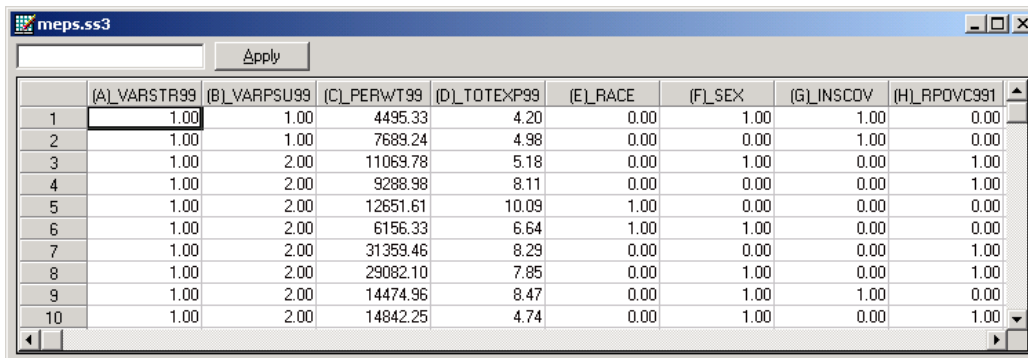
Contents

<i>Three-level models for continuous outcomes</i>	1
The data.....	1
The models.....	3
Example: A random intercept model with 7 predictors.....	4
Example: A random intercept model with 3 predictors.....	12

Three-level models for continuous outcomes

The data

The data set used here forms part of the data library of the Medical Expenditure Panel Survey (MEPS). Collected in 1999, these data from a longitudinal national survey were used to obtain regional and national estimates of health care use and expenditure based on the health expenditures of a sample of U.S. civilian non-institutionalized participants. The survey sample design utilized stratification, clustering, multiple stages of selection, and disproportionate sampling. The sample was drawn from 143 strata, divided into 460 primary sampling units (PSUs). Information on 23,565 participants included positive person-level weights and forms the data set used here, excluding the 1,053 participants in the original data with zero person-level weights. Further exclusion, of respondents with no total expenditure reported, reduced the size of the data set to 19300. Data for the first 10 participants on most of the variables used in this section are shown below in the form of a SuperMix spreadsheet file, named **meps.ss3**, which is saved in the **Examples\Primer\Continuous** folder.



	(A)_VARSTR99	(B)_VARPSU99	(C)_PERWT99	(D)_TOTEXP99	(E)_RACE	(F)_SEX	(G)_INSCOV	(H)_RPOVC991
1	1.00	1.00	4495.33	4.20	0.00	1.00	1.00	0.00
2	1.00	1.00	7689.24	4.98	0.00	0.00	1.00	0.00
3	1.00	2.00	11069.78	5.18	0.00	1.00	0.00	1.00
4	1.00	2.00	9288.98	8.11	0.00	0.00	0.00	1.00
5	1.00	2.00	12651.61	10.09	1.00	0.00	0.00	0.00
6	1.00	2.00	6156.33	6.64	1.00	1.00	0.00	0.00
7	1.00	2.00	31359.46	8.29	0.00	0.00	0.00	1.00
8	1.00	2.00	29082.10	7.85	0.00	1.00	0.00	1.00
9	1.00	2.00	14474.96	8.47	0.00	1.00	1.00	0.00
10	1.00	2.00	14842.25	4.74	0.00	1.00	0.00	1.00

The variables of interest are:

- VARSTR99 is the stratum identification variable (143 strata in total).
- VARPSU99 is the PSU identification variable (460 PSUs in total).
- PERWT99F represents the final sample weight, with weights ranging between 307.16 and 80061.61, correcting for both non-response and adjustments to population control totals from the Current Population Survey.
- TOTEXP99 is the natural logarithm of the total health expenditure of a respondent in 1999, ranging between 0.69 and 12.24 and representing actual expenditure of between \$1.99 and \$206,721.
- RACE is an ethnicity indicator, with a value of 1 indicating white respondents, and 0 denoting all other ethnic groups as well as respondents for which ethnicity is not known. This variable was recoded from the original MEPS variable RACEX.
- SEX is a gender indicator, with a value of 0 indicating a male participant and 1 a female participant; recoded from the original MEPS variable RSEX.
- INSCOV is an indicator of the level of insurance coverage, where 0 indicates private coverage any time during 1999, and 1 indicates public coverage or no insurance at all during 1999.
- RPOVC991 to RPOVC995 are five indicator variables, each associated with a category of the original MEPS variable RPOVC99 which was constructed by dividing family income by the applicable poverty line (selection of which depended on family size and composition), expressed as a percentage.

Income is a variable that is often transformed using its natural logarithm. Doing so in effect causes the impact of each additional dollar to decrease as income increases. Logarithmic transformation is also useful in lessening the influence of outliers, as the natural logarithm of a variable is much less sensitive to extreme observations than is the variable itself.

The original MEPS variable RPOVC99 assumed a value of 1 for a family with "high" income level where family income was equal to or greater than 400% of the applicable poverty line, and a value of 2 for those with a "low income" level (associated with 125% to 200% of the poverty line). Families with "middle income", "near poor" and "negative or poor" levels of income relative to poverty line income were coded 3, 4, and 5 respectively. For the "middle income" category, the ratio (as percentage) of family income to poverty line was 200% to less than 400%. In the case of "near poor" families, the percentages ranged between 100% and 125%, and for "negative or poor", the family income was less than 100% of the relevant poverty line. Thus, a value of 1 on the indicator variable RPOVC991 indicates a family with income at the "high" level. The variables RPOVC992, RPOVC993, and RPOVC994 are associated with the categories "low income", "middle income" and "near poor" respectively. A value of 1 on the variable RPOVC995 indicates a family with "negative or poor" income level.

Note that as each of the five indicator variables for categories of RPOVC99 is coded 1 if a participant responded in that category and 0 otherwise, only four of the five indicator variables can be used in a model where an intercept is included. Here, we opted to create them prior to analysis, as illustration of that feature is not relevant to the example at hand.

The models

A multilevel model does not make provision for the specification of design related variables such as stratum or PSU. Instead, these design variables are used to define the hierarchical structure of the data. In this example, the stratum identification variable VARSTR99 is used as the level-3 identifier and the PSU identification variable VARPSU99 serves to identify level-2 units (*i.e.*, PSUs) nested within a given stratum. We thus use the design variables to define a three-level hierarchical structure, with participants as level-1 observations nested within PSUs, in turn nested within strata. While not explicitly acknowledging the survey design or offering a conventional design effect estimate to measure the difference in estimates obtained when implementing this design compared to estimates obtained under a simple random sample, a multilevel model offers the advantage of estimating the variation in total health care expenditure within and between PSUs.

A general three-level model for a response variable y depending on a set of r predictors x_1, x_2, \dots, x_r can be written in the form

$$y_{ijk} = \mathbf{x}'_{(f)ijk} \boldsymbol{\beta} + \mathbf{x}'_{(3)ijk} \mathbf{v}_i + \mathbf{x}'_{(2)ijk} \mathbf{u}_{ij} + \mathbf{x}'_{(1)ijk} \mathbf{e}_{ijk}$$

where $i = 1, 2, \dots, N$ denotes the level-3 units, $j = 1, 2, \dots, n_i$ the level-2 units, and $k = 1, 2, \dots, n_{ij}$ the level-1 units. In this context, y_{ijk} represents the response of individual k , nested within level-2 unit j and level-3 unit i . The model shown here consists of a fixed and a random part. The fixed part of the model is represented by the vector product $\mathbf{x}'_{(f)ijk} \boldsymbol{\beta}$, where $\mathbf{x}'_{(f)ijk}$ is a typical row of the design matrix of the fixed part of the model with, as elements, a subset of the r predictors. The vector $\boldsymbol{\beta}$ contains the fixed, but unknown parameters to be estimated. The vector products $\mathbf{x}'_{(3)ijk} \mathbf{v}_i$, $\mathbf{x}'_{(2)ijk} \mathbf{u}_{ij}$, and $\mathbf{x}'_{(1)ijk} \mathbf{e}_{ijk}$ denote the random part of the model at levels 3, 2, and 1 respectively. For example, $\mathbf{x}'_{(3)ijk}$ represents a typical row of the design matrix of the random part at level 3, and \mathbf{v}_i the vector of random level-3 coefficients to be estimated. The products $\mathbf{x}'_{(2)ijk} \mathbf{u}_{ij}$ and $\mathbf{x}'_{(1)ijk} \mathbf{e}_{ijk}$ serve the same purpose at levels 2 and 1 respectively. It is assumed that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ are independently and identically distributed (i.i.d.) with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(3)}$. Similarly, $\mathbf{u}_{i1}, \mathbf{u}_{i2}, \dots, \mathbf{u}_{in_i}$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(2)}$, and $\mathbf{e}_{ij1}, \mathbf{e}_{ij2}, \dots, \mathbf{e}_{ijn_{ij}}$ are assumed i.i.d., with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_{(1)}$.

A random intercept model with 7 predictors

Within this hierarchical framework, the first model fitted to the data uses the participant's gender, ethnicity, type of health insurance coverage, and measure of income relative to poverty level to predict the total expenditure on health care in 1999, with expenditure transformed to the natural logarithm of the actual expenses incurred.

$$\begin{aligned} \text{TOTEXP99}_{ijk} = & \beta_0 + \beta_1 * \text{SEX}_{ijk} + \beta_2 * \text{RACE}_{ijk} + \beta_3 * \text{INSCOV}_{ijk} + \\ & \beta_4 * \text{RPOVC991}_{ijk} + \beta_5 * \text{RPOVC992}_{ijk} + \beta_6 * \text{RPOVC993}_{ijk} + \\ & \beta_7 * \text{RPOVC994}_{ijk} + v_{i0} + u_{ij0} + e_{ijk} \end{aligned}$$

where β_0 denotes the average expected total expenditure on health care in 1999, and $\beta_1, \beta_2, \dots, \beta_7$ indicate the estimated coefficients associated with the fixed part of the model which contains the predictor variables SEX, RACE, INSCOV and the indicator variables for categories of income relative to the poverty level. The random part of the model is represented by v_{i0} , u_{ij0} and e_{ijk} , which denote the variation in average total health related expenditure over strata, between PSUs (or, in other words, over PSUs nested within strata) and between participants at the lowest level of the hierarchy.

A random intercept model with 3 predictors

To illustrate model comparison, a simpler model was fitted to the same data. In the previous model, the estimated coefficients of only 2 of the 4 indicator variables representing a respondent's position relative to the poverty line were significant. In the second model, only the ethnicity, gender, and level of insurance coverage were included as explanatory variables. The simplified model is formulated as

$$\text{TOTEXP99}_{ijk} = \beta_0 + \beta_1 * \text{SEX}_{ijk} + \beta_2 * \text{RACE}_{ijk} + \beta_3 * \text{INSCOV}_{ijk} + v_{i0} + u_{ij0} + e_{ijk}$$

where β_0 denotes the average expected total expenditure on health care in 1999, and β_1, β_2 and β_3 indicate the estimated coefficients associated with the fixed part of the model which contains the predictor variables SEX, RACE, and INSCOV. The random part of the model is represented by v_{i0} , u_{ij0} and e_{ijk} as previously explained.

Example: A random intercept model with 7 predictors

Importing the data

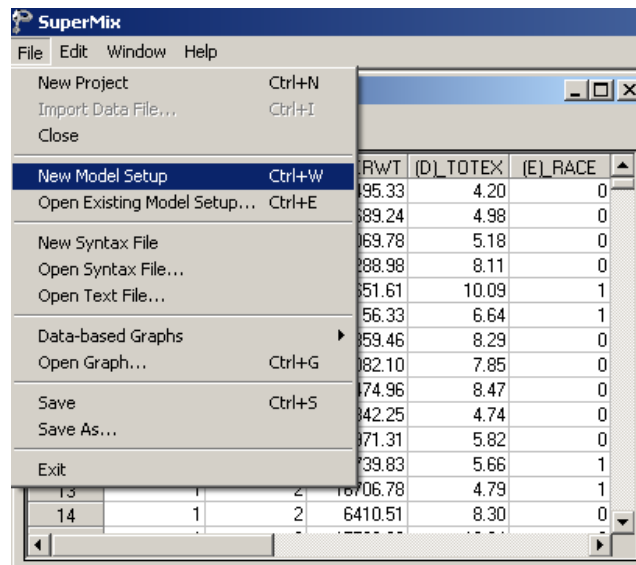
The model is fitted to the data in **meps.ss3**. The first step is to create the **ss3** file from an Excel spreadsheet named **meps.xls**. This is accomplished as follows:

- Use the **File, Import Data File** option to activate the display of an **Open** dialog box.
- Browse for the file **meps.xls** in the **Examples\Primer\Continuous** folder.
- Select the file and click the **Open** button to return to the main SuperMix window, where the contents of the Excel spreadsheet are displayed as the SuperMix system file with default name **meps.ss3**.

	(A)_VARSTR99	(B)_VARPSU99	(C)_PERWT99	(D)_TOTEXP99	(E)_RACE	(F)_SEX	(G)_INSCOV	(H)_RPOVC991
1	1.00	1.00	4495.33	4.20	0.00	1.00	1.00	0.00
2	1.00	1.00	7689.24	4.98	0.00	0.00	1.00	0.00
3	1.00	2.00	11069.78	5.18	0.00	1.00	0.00	1.00
4	1.00	2.00	9288.98	8.11	0.00	0.00	0.00	1.00
5	1.00	2.00	12651.61	10.09	1.00	0.00	0.00	0.00
6	1.00	2.00	6156.33	6.64	1.00	1.00	0.00	0.00
7	1.00	2.00	31359.46	8.29	0.00	0.00	0.00	1.00
8	1.00	2.00	29082.10	7.85	0.00	1.00	0.00	1.00
9	1.00	2.00	14474.96	8.47	0.00	1.00	1.00	0.00
10	1.00	2.00	14842.25	4.74	0.00	1.00	0.00	1.00

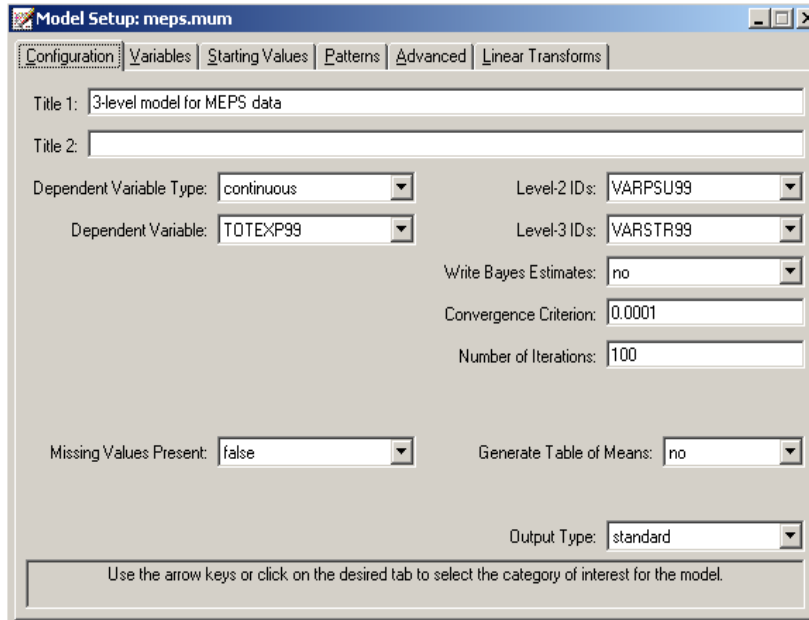
Setting up the analysis

The next step is to describe the model to be fitted. We use the SuperMix interface to provide the model specifications. From the main menu bar, select the **File, New Model Setup** option.

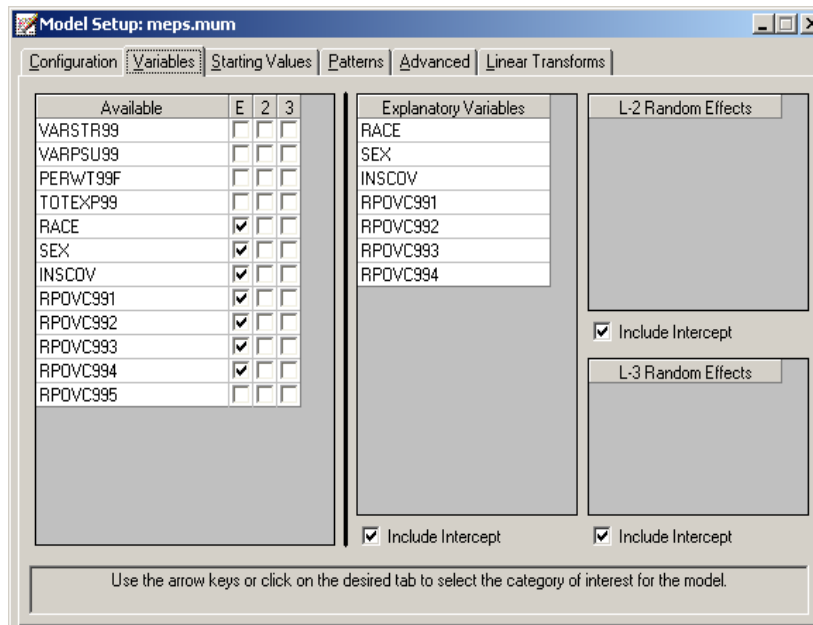


The **Model Setup** window that appears has six tabs: **Configuration**, **Variables**, **Starting Values**, **Patterns**, **Advanced**, and **Linear Transforms**. In this example, only the screens associated with the first two tabs are used.

As a first step, select the continuous outcome variable TOTEXP99 from the **Dependent Variable** drop-down list box. The stratum and cluster variables used to define the levels of the hierarchy are specified as Level-3 ID and Level-2 ID respectively by selecting them from the **Level-3 IDs** and **Level-2 IDs** drop-down list boxes. Enter a title for the analysis in the **Title 1** text box. In this example, default settings for all other options associated with the **Configuration** screen are used. Proceed to the **Variables** screen by clicking on this tab.



The **Variables** screen is used to specify the fixed and random effects to be included in the model. Start by selecting the explanatory (fixed) variables using the **E** check boxes next to the variables names in the **Available** grid at the left of the screen. Note that, as the variables are selected, the selected variables are listed in the **Explanatory Variables** grid. After selecting all the explanatory variables, the screen shown below is obtained. The **Include Intercept** check box in the **Explanatory Variables** grid is checked by default, indicating that an intercept term will automatically be included in the fixed part of the model.



The next step is to specify the random effects at levels 2 and 3 of the hierarchy. In this example, we want to fit a model with random intercepts at levels 2 and 3. By default, the **Include Intercept** check boxes in both the **L-2 Random Effects** and **L-3 Random effects** grids are checked. If these boxes are left checked, and no additional random effects are indicated using the **2** column in the **Available** grid to the left, the model fitted will be the random intercepts only model we intend to use. No further changes on this screen are necessary.

Before running the analysis, the model specifications have to be saved. Select the **File, Save As** option, and provide a name (**meps.mum**) for the model specification file. Run the analysis by selecting the **Run** option from the **Analysis** menu.

Discussion of results

Portions of the output file **meps.out** are shown below.

The screenshot shows a window titled "SuperMix - [meps.out]" with a menu bar (File, Analysis, Window, Help). The main content area displays the following text:

```

Numbers of observations
-----
Level 3 observations =      143
Level 2 observations =      460
Level 1 observations =    19300

LEVEL3 :      1      2      3      4      5      6      7      8
N2      :      2      2      2      2      2      3      2      2
N1      :     25     69     39     77    151     43     41     67

LEVEL3 :      9     10     11     12     13     14     15     16
N2      :      2      2      2     11      2      2      2      3
N1      :     97     21     56    336     54     58    141     53

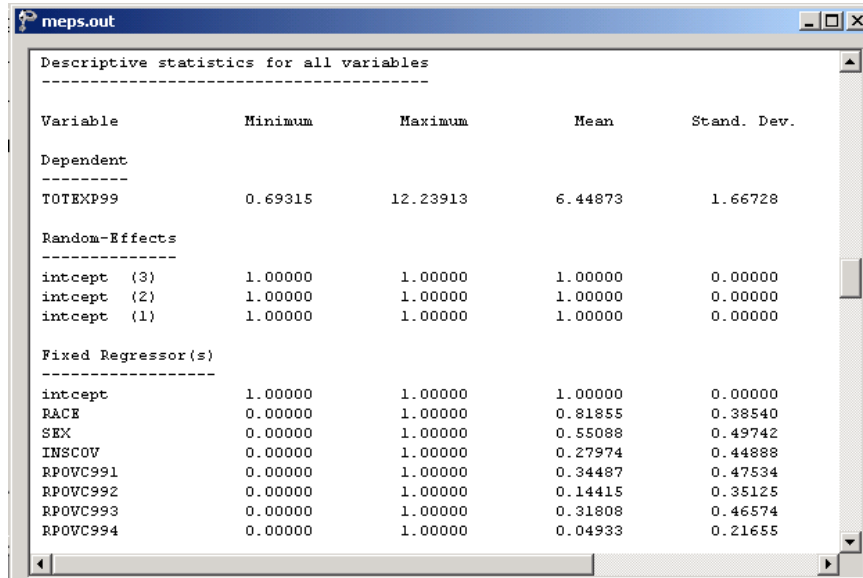
LEVEL3 :     17     18     19     20     21     22     23     24
N2      :      2      5      2      2      2      2      2      2
N1      :    293    198     61     42     38     50    138    120

LEVEL3 :     25     26     27     28     29     30     31     32
N2      :      3      2      2      2      2      2     10      2
N1      :    207     16     76     23     50     24    358     25

LEVEL3 :     33     34     35     36     37     38     39     40
N2      :      2      2      2      2     17      2      2      2
N1      :    303     36     51     72    512     82     27    155
  
```

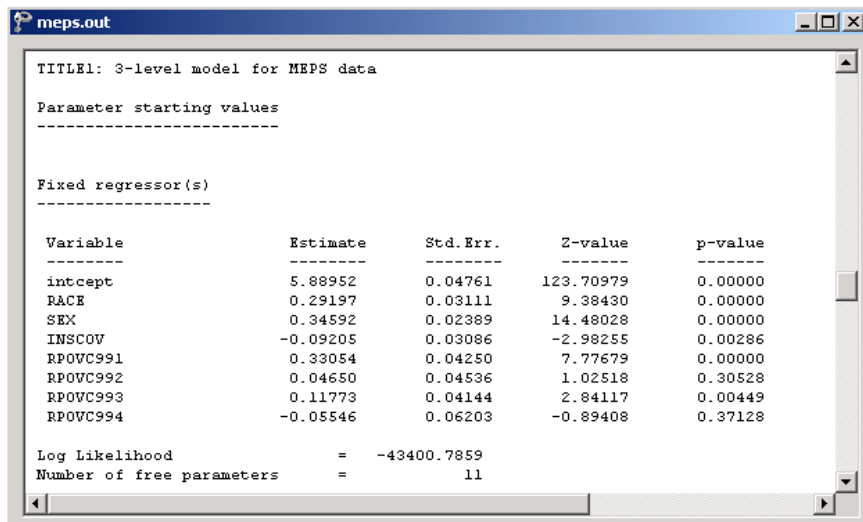
In the first section of the output file as shown above, a description of the hierarchical structure is provided. A total of 143 strata, 460 PSUs and information from 19,300 individual participants were included at levels 3, 2 and 1 of the model. This corresponds to the survey design described earlier. In addition, a summary of the number of PSUs and participants nested within each stratum is provided. For stratum number 1 (ID3: 1), data are available from only 25 participants nested within 2 primary sampling units (N2: 2). By contrast, for stratum number 12 (ID3: 12), data are available from 408 participants (N1: 336) nested within 11 primary sampling units (N2: 11).

The data summary is followed by descriptive statistics for all the variables included in the model.



Variable	Minimum	Maximum	Mean	Stand. Dev.
Dependent				
TOTEHP99	0.69315	12.23913	6.44873	1.66728
Random-Effects				
intcept (3)	1.00000	1.00000	1.00000	0.00000
intcept (2)	1.00000	1.00000	1.00000	0.00000
intcept (1)	1.00000	1.00000	1.00000	0.00000
Fixed Regressor(s)				
intcept	1.00000	1.00000	1.00000	0.00000
RACE	0.00000	1.00000	0.81855	0.38540
SEX	0.00000	1.00000	0.55088	0.49742
INSCOV	0.00000	1.00000	0.27974	0.44888
RPOVC991	0.00000	1.00000	0.34487	0.47534
RPOVC992	0.00000	1.00000	0.14415	0.35125
RPOVC993	0.00000	1.00000	0.31808	0.46574
RPOVC994	0.00000	1.00000	0.04933	0.21655

Descriptive statistics are followed by the starting values of the parameters that were used in the initial step of the iterative algorithm.



Variable	Estimate	Std. Err.	Z-value	p-value
intcept	5.88952	0.04761	123.70979	0.00000
RACE	0.29197	0.03111	9.38430	0.00000
SEX	0.34592	0.02389	14.48028	0.00000
INSCOV	-0.09205	0.03086	-2.98255	0.00286
RPOVC991	0.33054	0.04250	7.77679	0.00000
RPOVC992	0.04650	0.04536	1.02518	0.30528
RPOVC993	0.11773	0.04144	2.84117	0.00449
RPOVC994	-0.05546	0.06203	-0.89408	0.37128

Log Likelihood = -43400.7859
Number of free parameters = 11

The output describing the estimated **fixed effects** after convergence is shown next. The estimates are shown in the column with heading Estimate and correspond to the coefficients $\beta_0, \beta_2, \dots, \beta_7$ in the model specification. From the z-values and associated exceedance probabilities, we see that the coefficients associated with gender, ethnicity and insurance coverage type were all highly significant. Recall that a value of 1 for the ethnicity indicator variable RACE indicated that a participant was white, with a value of 0 assigned to participants from all other ethnic groups. The positive estimated coefficient for this variable indicates an increase of 0.3345 units in the logarithm of total health expenditure, holding all other predictors constant.

Similarly, female participants (coded "1" on the gender indicator SEX), are expected to have a total health expenditure 0.3469 higher than male participants if all other variables are held constant. In contrast, participants with public coverage or no coverage have a lower expected total expenditure, as indicated by the negative estimated coefficient -0.0598 .

```

Maximum likelihood estimates
-----
Fixed regressor(s)
-----
Variable          Estimate      Std. Err.    Z-value      p-value
-----
intcept           5.94083      0.05326     111.54352    0.00000
RACE              0.33446      0.03303     10.12702     0.00000
SEX              0.34691      0.02359     14.70287     0.00000
INSCOV           -0.05976     0.03108     -1.92277     0.05451
RPOVC991         0.25764      0.04320      5.96385     0.00000
RPOVC992         0.01777      0.04551      0.39037     0.69626
RPOVC993         0.07608      0.04175      1.82223     0.06842
RPOVC994        -0.06244     0.06211     -1.00538     0.31471

Log Likelihood      = -36869.7492
-2 Log Likelihood (Deviance) = 73739.4984
Akaike's Information Criterion = 73761.4984
Schwarz's Bayesian Criterion = 73794.0897
Number of free parameters = 11
  
```

Turning to the indicator variables associated with income relative to the poverty line, it can be seen that only one of the indicator variables, RPOVC991, has an estimated coefficients that is significantly different from zero at a 5% level of significance. In the case of families with a "high" income, the estimate of 0.2576 for RPOVC991 indicates an expected increase in expenditure, while for "near poor" families, the estimate of -0.0624 indicates an expected decrease in expenditure, holding all other variables constant.

In addition to the likelihood function value at convergence, a number of related statistical measures for model adequacy are available. The most common of these are the likelihood ratio test and Akaike's and Schwarz's criteria. Both the Akaike information criterion (AIC) and the Schwarz Bayesian criterion (SBC) are functions of the number of estimated parameters, and therefore "penalize" models with large numbers of parameters. In the SuperMix output file, all three of these are reported.

- The deviance is defined as $-2 \ln L$. For a pair of nested models, the difference in $-2 \ln L$ values has a χ^2 distribution, with degrees of freedom equal to the difference in number of parameters estimated in the models compared.
- The AIC was originally proposed for time-series models, but is also used in regression. It is defined as $-2 \ln L + 2r$, where r denotes the number of parameters estimated in the model. The model with minimum AIC, in a set of nested models, will be the most parsimonious according to this criterion.
- The SBC is defined as $-2 \ln L + r \log n$, where n denotes the number of units at the highest level of the hierarchy. A smaller value of this criterion would indicate the most parsimonious of the models being compared.

Estimated outcomes for different groups

To evaluate the expected effect of the measure of a family's income on the corresponding projected expenditure, suppose that the variables RACE, SEX, and INSCOV are held at zero, as would be the case for a nonwhite male participant with private insurance coverage. If such a participant originates from a family with "high" income, the logarithm of total health expenditure is expected to be

$$\begin{aligned} & \hat{\beta}_0 + \hat{\beta}_4 (\text{RPOVC991})_{ij} + \hat{\beta}_5 (\text{RPOVC992})_{ij} + \hat{\beta}_6 (\text{RPOVC993})_{ij} + \hat{\beta}_7 (\text{RPOVC994})_{ij} \\ &= \hat{\beta}_0 + \hat{\beta}_4 \\ &= 5.94083 + 0.25764 \\ &= 6.19847 \end{aligned}$$

which translates to a projected total expenditure of $e^{6.19847} = \$492$. In contrast, for a participant with similar demographic background and coverage from a "near poor" family, we obtain a projected total expenditure of

$$\begin{aligned} & e^{\hat{\beta}_0 + \hat{\beta}_7} \\ &= e^{5.94083 - 0.06244} \\ &= \$357.23. \end{aligned}$$

The predicted total expenditure (as natural logarithm) for similar participants from "low", "middle" or "negative or poor" families are similarly obtained by calculating $e^{\hat{\beta}_0 + \hat{\beta}_5}$, $e^{\hat{\beta}_0 + \hat{\beta}_6}$ and $e^{\hat{\beta}_0}$ respectively.

In Table 3.1, the predicted total health expenditure is given for respondents with high or near poor family income, for each of the subpopulations formed by gender, ethnicity, and insurance coverage. For purposes of the comparison, results are expressed in U.S. dollars, rather than in the natural logarithmic units of the outcome variable TOTEXP99. Respondents from families with high income consistently outspend their near poor counterparts by at least 100%, regardless of gender, ethnicity, or level of insurance coverage. In families with high income, female respondents spent more in 1999 than their male counterparts, regardless of ethnicity. This is generally also true for near poor respondents. It is also apparent that the total health expenditure in 1999 was higher for respondents with private insurance than for respondents with public or no coverage, and that white respondents spent more than respondents from other ethnic groups, regardless of gender or the level of family income. From exploratory analyses, we know that the outcome variable TOTEXP99, when expressed in terms of dollar values instead of natural logarithmic units, is highly skewed. It has a median of \$615 and a mean \$2,492. When this is taken in account, we can conclude that, generally speaking, white females spent more on health in 1999 than 50% of all respondents in the sample.

Table 3.1: Predicted total health expenditure for various subgroups

Group	Male (SEX = 0)		Female (SEX = 1)	
	Insurance coverage:		Insurance coverage:	
	Private (INSCOV=0)	Public/none (INSCOV = 1)	Private (INSCOV=0)	Public/none (INSCOV = 1)
Respondents with high family income (RPOVC991 = 1)				
Nonwhite (RACE = 0)	\$492	\$463	\$696	\$656
White (RACE = 1)	\$687	\$648	\$972	\$916
Respondents with near poor income (RPOVC994 = 1)				
Nonwhite (RACE = 0)	\$357	\$337	\$505	\$476
White (RACE = 1)	\$499	\$470	\$706	\$665

The output for the **random part** of the model follows, and is shown in the image below. There is significant variation in the average estimated total health expenditure at all levels, with the most variation over the participants (level 1), and the least variation over strata (level-3).

```

Variance/covariance components
-----
Level 3
-----
intcept /intcept      Estimate      Std. Err.      Z-value      p-value
-----
                      0.03494      0.01055      3.31114      0.00093

Level 2
-----
intcept /intcept      Estimate      Std. Err.      Z-value      p-value
-----
                      0.05032      0.00968      5.19828      0.00000

Level 1
-----
intcept /intcept      Estimate      Std. Err.      Z-value      p-value
-----
                      2.62733      0.02704      97.18098     0.00000

```

An estimate of the level-2 cluster effect, for example, is obtained as

$$\frac{0.05032}{0.03494 + 0.05032 + 2.62733} \times 100\% = 1.86\%$$

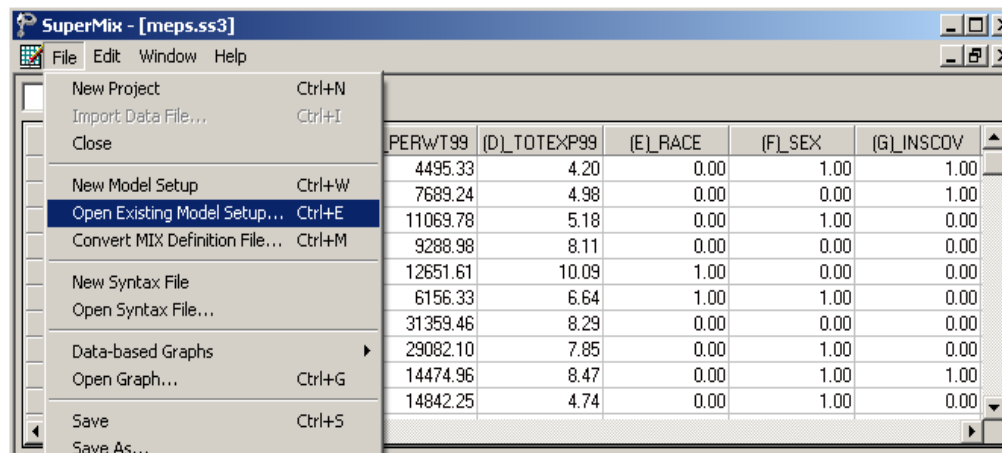
indicating that only 1.86% of the total variance in expenditure explained is at level 2 of the model.

Example: A random intercept model with 3 predictors

To illustrate model comparison, a simpler model was fitted to the same data. In the previous model, the estimated coefficients of only 2 of the 4 indicator variables representing a respondent's position relative to the poverty line were significant. In this model, only the ethnicity, gender, and level of insurance coverage were included as explanatory variables.

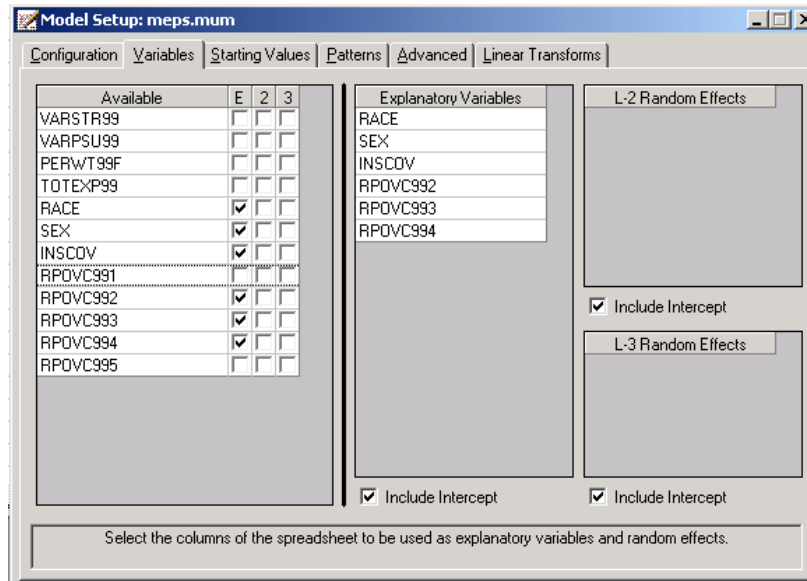
Setting up the analysis

The model specification file from the previous example can be used as the basis for the simplified model. As a previously saved model specification file can only be opened once the associated **ss3** file is open, start by using the **File, Open** option to browse for and open **meps.ss3**. Next, open the model specification file **meps.mum** by selecting the **Open Existing Model Setup** option from the **File** menu.



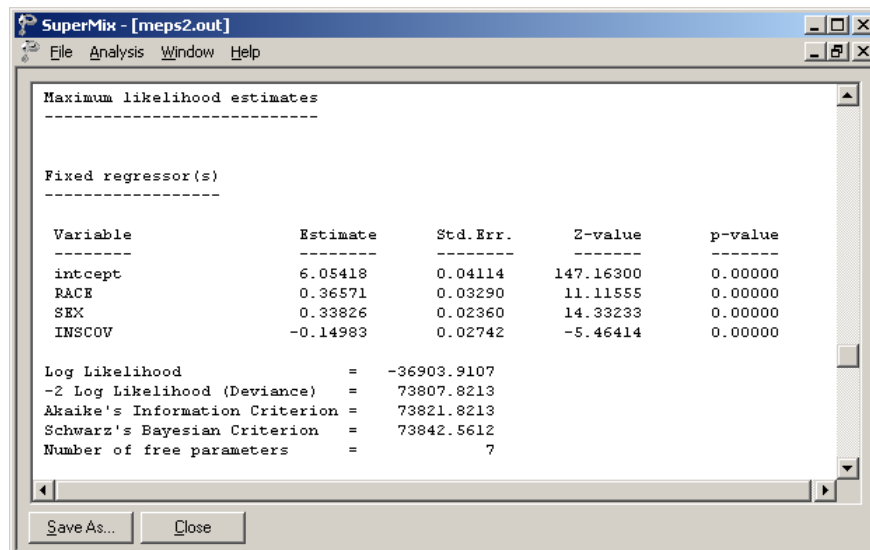
Once the **Model Setup** window is displayed, click on the **Variables** tab. Delete the first of the four indicator variables (**RPOVC991**) by unchecking the check box next to this variable as shown below.

Remove the other three indicator variables in the same way, and save the revised model specification file as **meps2.mum** using the **File, Save As** option. Finally, click the **Analysis, Run** option to start the analysis.



Discussion of results

A portion of the output file **meps2.out** is shown below. The estimates of the fixed effects are close to those obtained for the previous model.



Recall that the fit measures obtained for the previous model were:

Log Likelihood = -36869.7492
 Deviance = 73739.4984
 AIC = 73761.4984
 SBC = 73784.0897
 Number of free parameters = 11

All of the reported statistical measures for model adequacy indicate that the previous model, where the variables RPOVC991 to RPOVC994 were included, offers a better fit to the data.