

階層的線形モデルによるデータの分析例

－ ソフトウェアの使い方を中心に －

奥村太一*

東京大学大学院教育学研究科 博士課程

2006 年 5 月 26 日

1 はじめに

階層的線形モデルは、集団単位で集められたデータや個人内反復測定データなど観測値相互の独立性が確保できないようなデータを分析するのに極めて有用な統計モデルである。これまで、階層的線形モデルを用いてデータの分析を行うことのできるソフトウェアがいくつか提供されてきた。

今回は、そうしたソフトウェアのうち最もポピュラーなものとして、S. W. Raudenbush らによって開発された HLM6 の student edition を用いて、実際のデータを用いたデータ解析の実習を行う。また、それ以外に利用できるパッケージとして、SAS/STAT の MIXED プロシジャ、およびフリーの統計解析パッケージとして最近注目されている R を用いた解析例を簡単に紹介する。

2 HLM6 について

HLM は、階層的線形モデルを分析するためのソフトウェアとしてその理論的整備に大きく貢献した A. S. Bryk や S. W. Raudenbush らによって開発されたものである。現在、HLM は SSI 社^{*1}から提供されているが、最新バージョンである HLM6 では student edition が無料で公開されている^{*2}。student edition は、商用版に比べて扱えるサンプルサイズやモデルの大きさに制約があるものの、分析手法に関しては商用版と同じ機能を有する。また、多くのサンプルデータが用意されているため、それらを使ってさまざまな分析の練習を行うことも可能である。

これから具体的に見るように、HLM6 は、(1) 優れた GUI 機能を有すること、(2) そのため比較的直感的に操作を行えること、(3) 豊富なグラフィック機能を有することなどから、統計を専門にしないユーザーにとって最も使い勝手のよいソフトウェアの一つであると考えられる。

ソフトウェアの名称である HLM とは、*Hierarchical Linear Models* の略である。

* okumurin@p.u-tokyo.ac.jp

*1 <http://www.ssicentral.com/home.htm>

*2 <http://www.ssicentral.com/hlm/student.html> からダウンロード可能。

3 HLM6 による分析の手順

3.1 今回の分析例

今回は、最も単純な 2 水準 (2 レベル) のモデルの分析例を示すために、High School and Beyond 調査からのデータを利用する。このデータは HLM6 をダウンロードすれば、サンプルデータとして提供されるものである。このデータは、学校 → 生徒という 2 段階のサンプリングを経て得られたものであり、生徒個人の情報として数学の成績 (MATHACH)、また学校に関する情報として学校の区分 (school SECTOR: カトリックか公立か) といった変数が含まれている。また、各生徒について家庭の社会経済的地位を表す尺度の得点 (Socio-Economic Status: SES) が測定されている他、それを各学校ごとに平均した値がその学校に通う生徒の平均的な社会経済的地位を表す変数 (MEANSES) として計算されている。

今回は、生徒の数学の成績 (MATHACH) が個々の生徒の家庭の社会経済的地位 (SES)、および生徒が属する学校の区分 (SECTOR)、そして学校に通う生徒の平均的な社会経済的地位 (MEANSES) からどのように説明されるかを 2 レベルの階層的線形モデルによってモデル化し、実際に分析を行う。

表 1 今回の分析で使用する変数

レベル 1 (生徒)	レベル 2 (学校)
MATHACH (数学の成績)	SECTOR (学校の区分)
SES (家庭の社会経済的地位)	MEANSES (SES の学校平均)

仮に HLM6 を C:\Program Files\HLM6S にインストールした場合、C:\Program Files\HLM6S\Examples\AppendxA にデータファイルが含まれる。

同じデータを使った同様の分析例が Raudenbush and Bryk (2002) に載っている。

モデルについて

レベル 1: ユニットは生徒 (7185 名) である。このレベルでは、生徒の社会経済的地位 (SES) を用いて数学の成績 (MATHACH) を予測する回帰式を立てる。

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij} - \bar{SES}_{.j}) + r_{ij} \quad (1)$$

レベル 2: ユニットは学校 (160 校) である。このレベルでは、レベル 1 の回帰式における学校ごとの切片と傾きを学校の区分 (SECTOR) および学校の平均的な社会経済的地位 (MEANSES) によって予測する回帰式を立てる。

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SECTOR_j) + \gamma_{02}(MEANSES_j) + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(SECTOR_j) + \gamma_{12}(MEANSES_j) + u_{1j} \quad (3)$$

ここで、 i は生徒、 j は学校を表す。また、 $\bar{SES}_{.j}$ は、 j 番目の学校の生徒の SES の平均である。このようなセンタリングを行う方法および理由については後で述べる。

$$SECTOR = \begin{cases} 1 & \text{(Catholic)} \\ 0 & \text{(public)} \end{cases}$$

レベル 2 の式には確率的に変動する誤差成分 (u_{0j} , u_{1j}) が含まれている。これは、レベル 1 の切片および傾きには今回用意した 2 つの説明変数 (SECTOR, MEANSES) のみでは説明できない独自の成分があるということを示している。

3.2 分析の手順

3.2 分析の手順

HLM6 では、以下のような手順で分析を行う。また、分析を行う過程で様々なファイルが作成される。

1. データファイルの作成
2. データファイルを読み込み、十分統計量を計算する
 - 分析に用いるモデルの選択（レベルの数など）
 - データファイルの読み込み、変数の指定など
 - MDMT ファイル（.mdmt）を作成（MDM ファイルで指定した内容を格納）
 - 十分統計量を計算し、MDM ファイル（.mdm）として保存
 - STS ファイル（.sts）を確認（基本統計量を格納）
3. モデル式を指定し、出力や推定に関するオプションを指定
 - 指定した内容を、HLM ファイル（.hlm）もしくは MLM ファイル（.mlm）で保存
4. 分析の実行
5. 結果の確認

3.3 データファイルの作成

HLM6 においては、データファイルはそれぞれのレベルごとに 1 つずつ作成する必要がある。すなわち 2 レベルのモデルであれば 2 種類の、3 レベルのモデルであれば 3 種類のデータファイルを作成することになる。

HLM6 においてデータファイルを作成するには、以下の 2 つの方法がある。

1. ASCII 形式（通常のテキストファイル）を用いる方法
2. 他の統計パッケージのファイルを入力として用いる方法

以下、各々について説明する。

今回の分析例では 2 レベルのモデルを用いるため、必要となるデータファイルの数は 2 つである。

SPSS, SAS, SYSTAT, STATA で作成したデータファイルが使える。

3.3.1 ASCII 形式によるデータファイルの作成

データのフォーマット

HLM6 では、データ入力のフォーマットが厳格に定められている。以下にその規則のうち重要と思われるものを挙げる。

- データの読み込みはカラム指定で行われるので、スペースを調整するなどして各変数ごとに縦の列をそろえておく。
- データファイルには ID を表す変数を入れる必要がある。ID 変数は、自身よりも一段上のレベルのケースとの対応を示すものであり、そのレベルのケースの ID ではない。

今回の分析例では、各生徒の所属する学校の ID である。生徒個人の ID ではないことに注意。

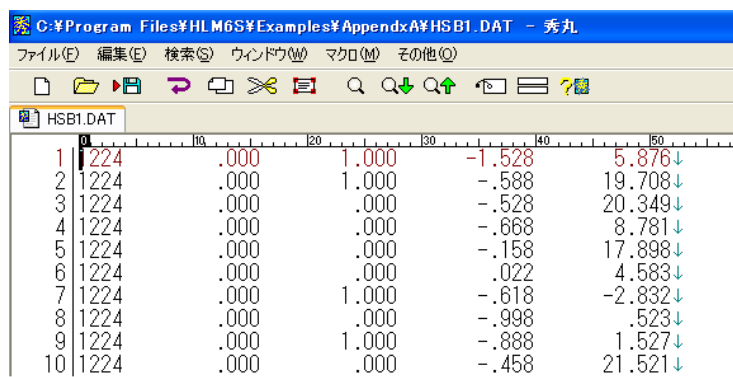
3.3 データファイルの作成

- すべてのデータは ID 変数の値ごとにまとめられていなければならない。
- 同じファイル内の ID 変数の桁はそろえておかななくてはならない。
- 欠測値は、レベル 1 のデータファイルでのみ許される。

同じ学校に所属する生徒はまとめて同じところに。

欠測値は”.”で代用できるが、欠測値のあるケースは分析の途中で除外される。(8 ページを参照。)

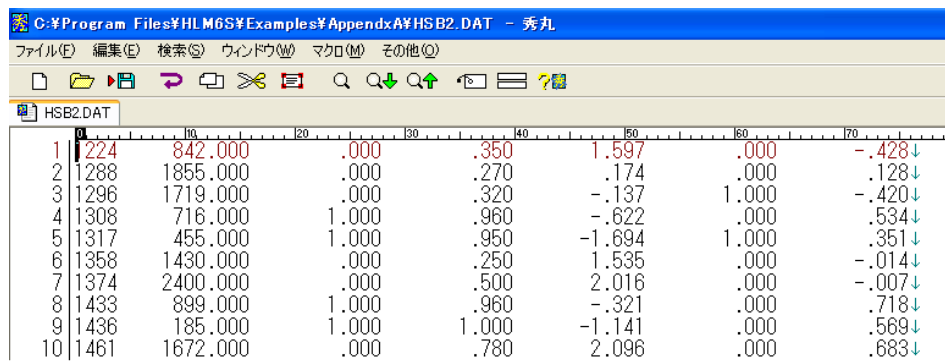
レベル 1 のデータファイル (HSB1.DAT)



	ID	SES	MATHACH
1	1224	.000	1.000
2	1224	.000	1.000
3	1224	.000	.000
4	1224	.000	.000
5	1224	.000	.000
6	1224	.000	.000
7	1224	.000	1.000
8	1224	.000	.000
9	1224	.000	1.000
10	1224	.000	.000

↑ ID ↑ SES ↑ MATHACH

レベル 2 のデータファイル (HSB2.DAT)



	ID	SECTOR	MEANSES
1	1224	.000	.350
2	1288	.000	.270
3	1296	.000	.320
4	1308	1.000	.960
5	1317	1.000	.950
6	1358	.000	.250
7	1374	.000	.500
8	1433	1.000	.960
9	1436	1.000	1.000
10	1461	.000	.780

↑ ID ↑ SECTOR MEANSES ↑


3.3.2 SPSS ファイルによるデータセットの作成

SPSS のデータファイル形式 (.sav) によってデータセットを作成することもできる。この方法では各セルに 1 つずつデータを入れることから、ASCII のようにカラムをそろえるなどの配慮を行わなくてよく、後で説明する MDM ファイルの作成も ASCII 形式のデータファイルを用いた場合よりも簡単にできるなど、何かと便利である。

その他の規則は ASCII による方法のところでも述べたのと同様である。

3.4 MDM ファイルの作成

レベル 1 のデータファイル (HSB1.SAV)



	id	minority	female	ses	mathach	v
1	1224	.000	1.000	-1.528	5.876	
2	1224	.000	1.000	-.588	19.708	
3	1224	.000	.000	-.528	20.349	
4	1224	.000	.000	-.668	8.781	
5	1224	.000	.000	-.158	17.898	
6	1224	.000	.000	.022	4.583	
7	1224	.000	1.000	-.618	-2.832	
8	1224	.000	.000	-.998	.523	
9	1224	.000	1.000	-.888	1.527	
10	1224	.000	.000	-.458	21.521	

レベル 2 のデータファイル (HSB2.SAV)



	id	size	sector	pracad	disclim	himinty	meanses
1	1224	842.000	.000	.350	1.597	.000	-.428
2	1288	1855.00	.000	.270	.174	.000	.128
3	1296	1719.00	.000	.320	-.137	1.000	-.420
4	1308	716.000	1.000	.960	-.622	.000	.534
5	1317	455.000	1.000	.950	-1.694	1.000	.351
6	1358	1430.00	.000	.250	1.535	.000	-.014
7	1374	2400.00	.000	.500	2.016	.000	-.007
8	1433	899.000	1.000	.960	-.321	.000	.718
9	1436	185.000	1.000	1.000	-1.141	.000	.569
10	1461	1672.00	.000	.780	2.096	.000	.683

プログラムとともに提供されているサンプルデータでは、小数点以下の桁を表示しないよう設定されているはずであるが、ここではわかりやすいように ASCII 形式のデータと同じように表示されるよう設定を変えてある。データの内容そのものは同じである。

3.4 MDM ファイルの作成

HLM6 では、実際の分析は生データを直接用いるのではなく、それをいったん処理した MDM (Multivariate Data Matrix) ファイル (.mdm) の情報を用いて行われる。従って、分析を行う前にまず生データを入力したファイルから MDM ファイルを作成する必要がある。また、いったん MDM ファイルを作っておけば、その後の分析には生データファイルを直接用いることなく、すでに作った MDM ファイルを繰り返して使うことができる。

ここでは、前項で取り上げた ASCII 形式で作成されたデータファイル (.dat ファイル) および SPSS によって作成されたデータファイル (.sav ファイル) から MDM ファイルを作成する方法について解説する。

MDM ファイルに相当するものは、以前のバージョンである HLM5 までは、SSM (Sufficient Statistics Matrix) ファイルといっていたが、バージョン 6 への改定の際に変更された。

3.4 MDM ファイルの作成

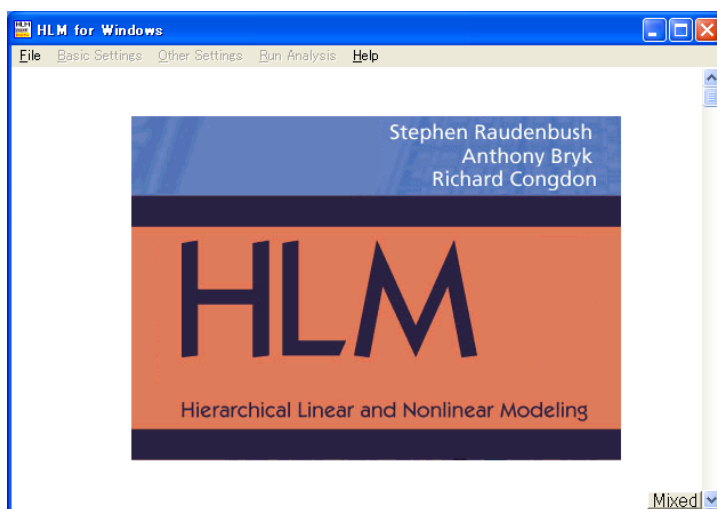
3.4.1 HLM6 の起動

まず、HLM6 のショートカットをダブルクリックして HLM6 を起動する。



もしショートカットがなければ、"HLM6S"フォルダ内の"WHLMS.exe"ファイルをダブルクリックしてもよい。あるいは、"WHLMS.exe"を右クリックし「送る」[デスクトップ(ショートカットを作成)]で、デスクトップ上にショートカットを作成することができる。

すると、以下のようなウィンドウが立ち上がるはずである。HLM6 の操作は基本的にこのウィンドウから行うことになる。

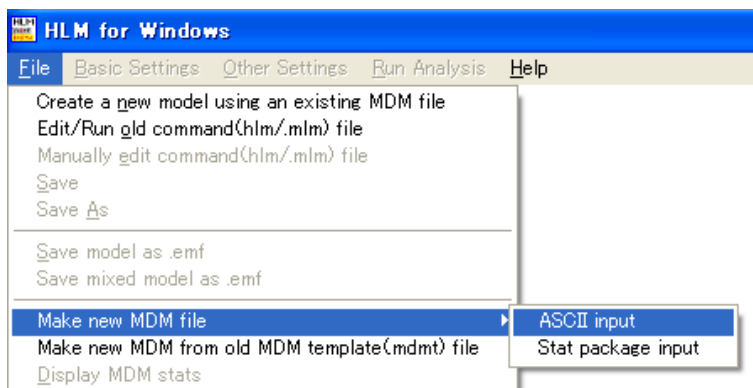


3.4.2 ASCII 形式によるデータファイルを用いた MDM ファイルの作成

ここでは、ASCII 形式で作成したデータファイルを用いて MDM ファイルを作成する方法について説明する。

前項で、データファイルはモデルの水準の数だけ必要であると説明したが、MDM ファイルは 1 つのモデルに 1 つ作ればよい。

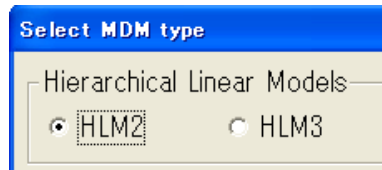
まず、[File]-[Make new MDM file]-[ASCII input] とメニューを選択する。



3.4 MDM ファイルの作成

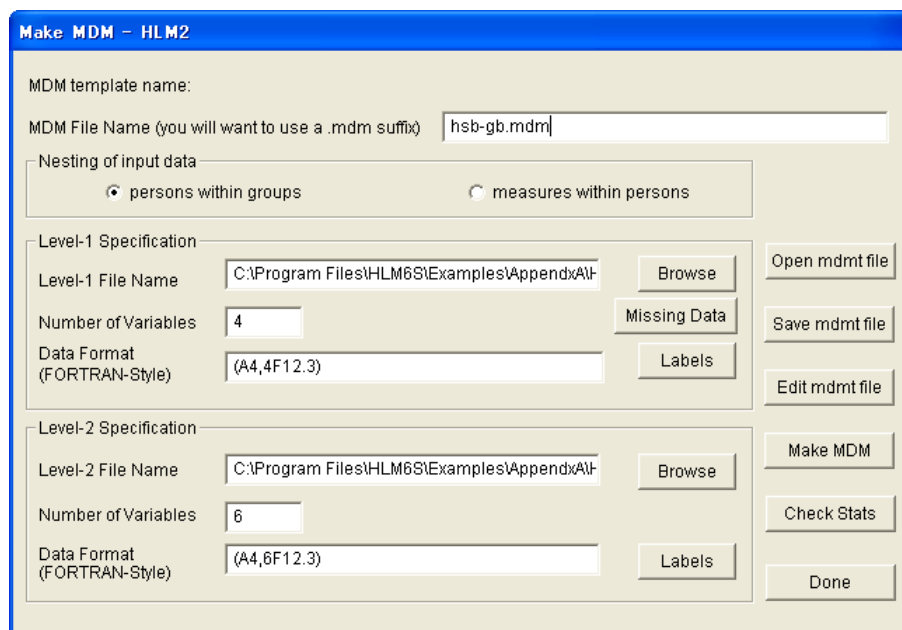
次に、分析するモデルのタイプを選択し、[OK] をクリックする。

今回の例では 2 レベルモデルなので、“HLM2”を選択する。



“Make MDM - HLM2”が開く。ここで、データファイルから MDM ファイルを作成するためのデータに関する情報を指定することになる。

以下に、各項目について入力・指定の方法について述べる。



MDM File Name

作成する MDM ファイルの名前を拡張子つきで入力する。

ここでは、“hsb-gd.mdm”とした。

Nesting of input data

データの種類が、「集団 - 個人」であるか「個人 - 測定」であるかを指定する。分析の内容には直接関係ないが、後にモデル式の指定をする際にどちらを選択したかで記号の区別がなされることになる。

ここでは、「学校という集団内の個人」を対象にしたデータであることから“persons within groups”を選択した。

Level-1 Specification

ここでは、レベル 1 のデータセットについての情報を指定することになる。

Level-1 File Name: レベル 1 のデータファイルをフルパスで入力する。あるいは、右にある“Browse”ボタンをクリックしてファイルを選択すれば、自動的にそのファイルのフルパスが入力される。

ここでは“C:\Program Files\HLM6S\Examples\AppendxA\hsb1.dat”と入力するか、そのファイルを選択する。

3.4 MDM ファイルの作成

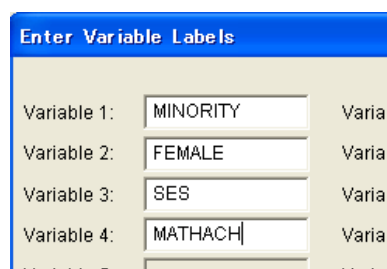
Number of Variables: ID 変数を除いた変数の数を半角で入力する。

Missing Data: データに欠側がある場合はここで宣言しておく。欠測値があれば、欠側のあるケースをどの段階で削除するか（MDM ファイルを作る段階/分析を行う段階）を選択することになる。

Data Format: データのフォーマット（変数のカラム）の指定を行う。その際の規則は以下のとおりである。

- "A" は ID 変数を読み込むための記述子である。"A" の後に、ID 変数が何桁あるかを数字で入力する。
- "F" は実際のデータを読むための記述子である。"F" の前にある数字は繰り返し数（ID 変数を除く変数の数）である。"F" の後には、"[カラム数].[小数点以下の桁数]" を入力する。
- "X" は、カラムを飛ばすための記述子である。
- "/" は、データの読み込みが 2 行以上にわたる際に改行を指定する記述子である。
- 以上の各記述子は、カンマ（","）で区切る。

Labels: クリックしてウィンドウを開き、変数のラベル（変数名）を半角英数字で入力する。日本語で入力すると文字化けしてしまうので注意。



Level-2 Specification

レベル 2 のデータファイルに関する情報を入力する。基本的にレベル 1 の場合と同様に行えばよいが、欠測はレベル 2 以上では許容されないためこれに関しては指定を行わない。

Save mdmt file

[Save mdmt file] をクリックして、MDMT(MDM Template)ファイル(.mdmt)を作る。MDMT ファイルには、MDM ファイルを作るために指定した情報^{*3}が格納される^{*4}。

^{*3} データファイルの種類や変数の数、カラム指定、欠測値に関する情報など。実際にテキストエディタで開いてみると MDM ファイルを作る際に指定した内容が確認できる。

^{*4} MDMT ファイルを用いれば、以前に作成した MDM ファイルを簡単に編集し直すことができる。[File]-[Make new MDM from old MDM template (.mdmt) file] メニューから MDMT ファイルを指定すれば、MDM ファイルを作成する画面に移動する。

ここでは、"4" とした。
今回の例では欠測はないのでそのままでもよい。デフォルトでは欠測はないものとして扱われる。

ここでは"(A4,4F12.3)" と入力する。"(" や ")" を忘れると警告が出る。

"4F12.3" は、12 カラムを占める小数点以下 3 桁までのデータが 4 つあるという意味になる。実際にデータファイル"hsb1.dat"を確認されたい。

例えば"8X"であればデータ読み込みの際に 8 カラム飛ばすということになる。

ここでは、順に"MINORITY", "FEMALE", "SES", "MATH-ACH"とする。

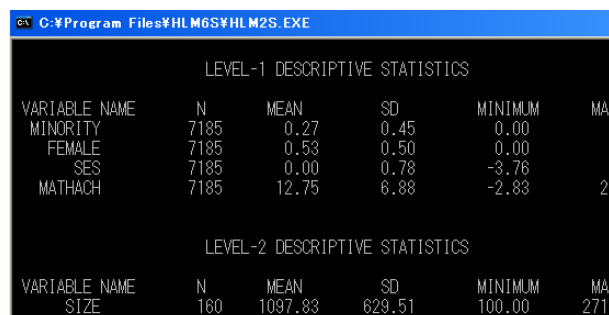
データファイルは、C:\¥Program Files¥HLM6S¥Examples¥AppendxA¥hsb2.dat と入力するか、そのファイルを選択する。変数の数は 6 個で、Data Format には"(A4,6F12.3)" と入力する。また、変数名については順に"SIZE", "SECTOR", "PRACAD", "DISCLIM", "HIMINTY", "MEANSES" とする。

3.4 MDM ファイルの作成

Make MDM

[Make MDM] ボタンをクリックして、MDM ファイルをつくる。次のような画面が現れて、MDM ファイルが作られたことがわかる。

この画面は、数秒間現れたのち自動的に消える。



The screenshot shows a window titled 'C:\Program Files\HLM6S\HLM2S.EXE'. It displays two tables of descriptive statistics. The first table, 'LEVEL-1 DESCRIPTIVE STATISTICS', has columns for VARIABLE NAME, N, MEAN, SD, MINIMUM, and MAXIMUM. The second table, 'LEVEL-2 DESCRIPTIVE STATISTICS', has the same columns. The data is as follows:

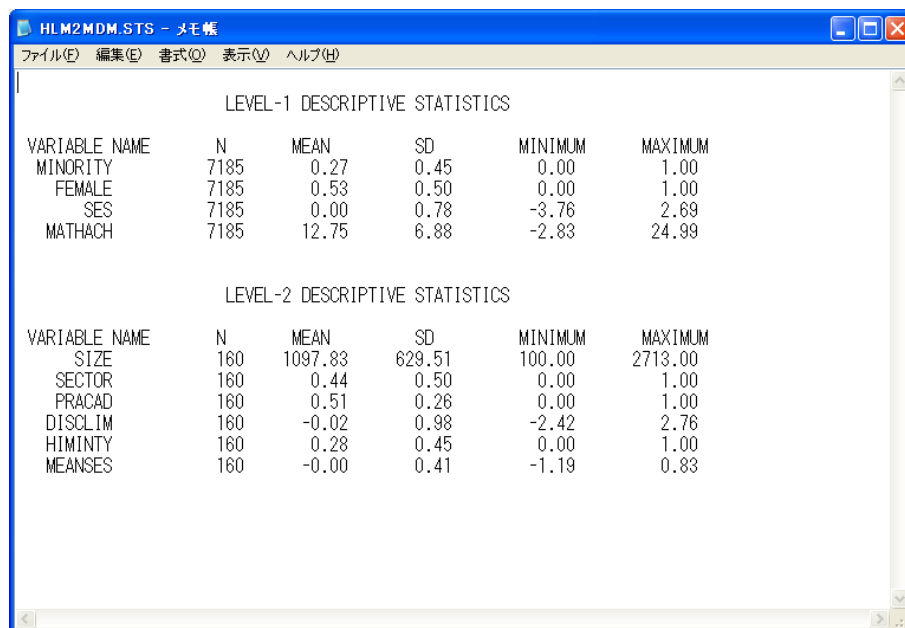
LEVEL-1 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MINORITY	7185	0.27	0.45	0.00	1.00
FEMALE	7185	0.53	0.50	0.00	1.00
SES	7185	0.00	0.78	-3.76	2.69
MATHACH	7185	12.75	6.88	-2.83	24.99

LEVEL-2 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SIZE	160	1097.83	629.51	100.00	2713.00
SECTOR	160	0.44	0.50	0.00	1.00
PRACAD	160	0.51	0.26	0.00	1.00
DISCLIM	160	-0.02	0.98	-2.42	2.76
HIMINTY	160	0.28	0.45	0.00	1.00
MEANSES	160	-0.00	0.41	-1.19	0.83

Check Stats

[Check Stats] ボタンをクリックすると、各変数の基本統計量を格納した STS ファイル (hlm2mdm.sts) を確認することができる。

STS ファイルは自動的に作成され、あらかじめファイル名を指定することはできない。



The screenshot shows a window titled 'HLM2MDM.STS - メモ帳'. It displays two tables of descriptive statistics, identical to the ones in the previous screenshot. The first table, 'LEVEL-1 DESCRIPTIVE STATISTICS', has columns for VARIABLE NAME, N, MEAN, SD, MINIMUM, and MAXIMUM. The second table, 'LEVEL-2 DESCRIPTIVE STATISTICS', has the same columns. The data is as follows:

LEVEL-1 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MINORITY	7185	0.27	0.45	0.00	1.00
FEMALE	7185	0.53	0.50	0.00	1.00
SES	7185	0.00	0.78	-3.76	2.69
MATHACH	7185	12.75	6.88	-2.83	24.99

LEVEL-2 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SIZE	160	1097.83	629.51	100.00	2713.00
SECTOR	160	0.44	0.50	0.00	1.00
PRACAD	160	0.51	0.26	0.00	1.00
DISCLIM	160	-0.02	0.98	-2.42	2.76
HIMINTY	160	0.28	0.45	0.00	1.00
MEANSES	160	-0.00	0.41	-1.19	0.83

Done

[Done] ボタンをクリックすることで MDM ファイルの作成段階は終了し、モデルの指定を行う画面が現れる。

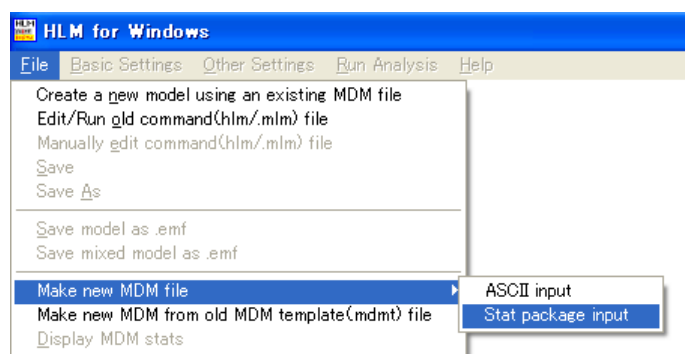
MDM ファイルが指定された後は、[File]-[Display MDM stats] で MDM ファイルに含まれる情報の一部を確認することができる。

3.4 MDM ファイルの作成

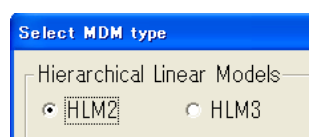
3.4.3 SPSS によるデータファイルを用いた MDM ファイルの作成

ここでは、SPSS によって作成した.sav 形式によるデータファイルを用いて MDM ファイルを作成する方法を解説する。

まず、[File]-[Make new MDM file]-[Stat package input] とメニューを選択する。

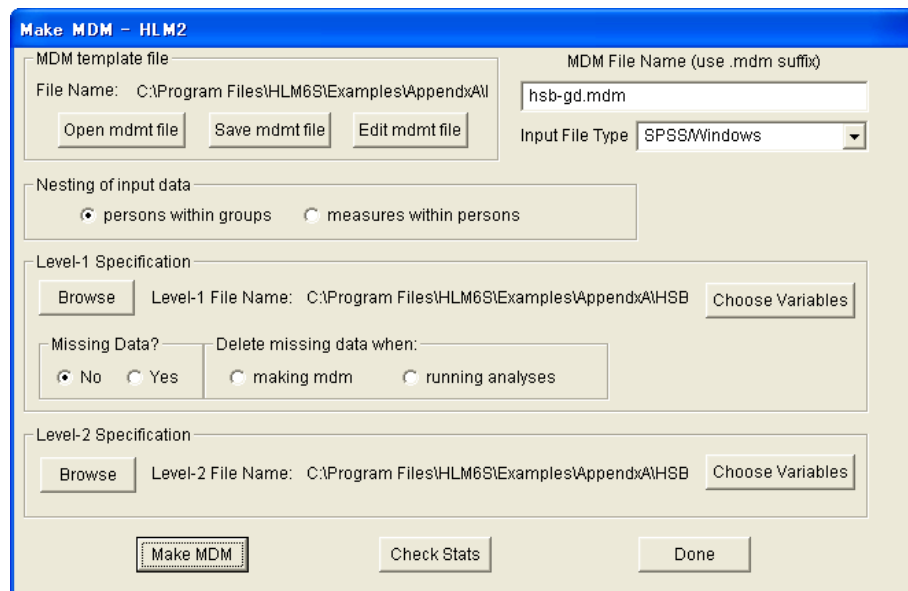


次に、分析するモデルのタイプを選択し、[OK] をクリックする。



ASCII 形式のときと同様、今回の例では 2 レベルモデルなので、「HLM2」を選択する。

「Make MDM - HLM2」が開く。



ここで、データファイルから MDM ファイルを作成するためのデータに関する情報を指定することになる。

以下に、各項目について入力・指定の方法について述べる。

3.4 MDM ファイルの作成

MDM File Name

作成する MDM ファイルの名前を拡張子 (.mdm) つきで入力する。

ここでは, "hsb-gd.mdm" とした。

Input File Type

ここで, データファイルの形式をプルダウンメニューから選択する。

ここでは"SPSS/Windows"を選択(デフォルト)。

Nesting of input data

データの種類の, 「集団 - 個人」であるか「個人 - 測定」であるかを指定する。

ここでは, 「学校という集団内の個人」を対象にしたデータであることから"persons within groups"を選択した。

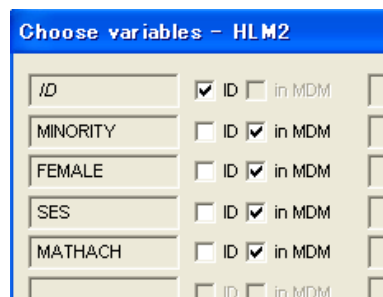
Level-1 Specification

ここでは, レベル 1 のデータファイルに関する指定を行う。

Browse: [Browse] ボタンをクリックしてファイルを選択し, "Level-1 File Name: " の右側に, データファイルのフルパスが表示されるのを確認。

ここでは"C:\Program Files\HLM6S\Examples\AppendxA\hsb1.sav" を選択した。

Choose Variables: [Choose Variables] をクリックして, レベル 1 のデータファイルから読み込む変数の指定 (四角をクリックしてチェックをつける) を行う。SPSS ファイルではすでに変数名がつけられているので, ここでは ID 変数および MDM ファイルに読み込む変数 (分析の対象としたい変数) を指定するだけでよい。



欠測値に関する指定: もし欠測値があれば "Missing Data?" で "Yes" を, なければ "No" を選択する。"Yes" を選択した場合は, 欠測値を含むケースをいずれの段階で除去するか (MDM ファイルを作る段階 / 分析を行う段階), 選択する。

hsb1.sav には欠測値はないのでデフォルトのままでよい。

Level-2 Specification

次に, レベル 2 のデータファイルに関する指定を行う。

Browse: [Browse] ボタンをクリックして, レベル 2 のデータファイルを指定する。

ここでは"C:\Program Files\HLM6S\Examples\AppendxA\hsb2.sav" を選択した。

Choose Variables: [Choose Variables] ボタンをクリックして, レベル 2 で読み込む変数の指定を行う。

3.4 MDM ファイルの作成



Save mdmt file

[Save mdmt file] をクリックして、MDM テンプレートファイルを保存する。

ここでは、”hsb-gd.mdmt”とした。

Make MDM

[Make MDM] をクリックして、MDM ファイルを作成する。次のような画面が現れて、MDM ファイルが作られたことがわかる。

この画面は、数秒間現れたのち自動的に消える。

LEVEL-1 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MINORITY	7185	0.27	0.45	0.00	1.00
FEMALE	7185	0.53	0.50	0.00	1.00
SES	7185	0.00	0.78	-3.76	2.69
MATHACH	7185	12.75	6.88	-2.83	24.99

LEVEL-2 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SIZE	160	1097.83	629.51	100.00	2713.00
SECTOR	160	0.44	0.50	0.00	1.00
PRACAD	160	0.51	0.26	0.00	1.00
DISCLIM	160	-0.02	0.98	-2.42	2.76
HIMINTY	160	0.28	0.45	0.00	1.00
MEANSES	160	-0.00	0.41	-1.19	0.83

Check Stats

[Check Stats] ボタンをクリックすると、各変数の基本統計量を格納した STS ファイル (hlm2mdm.sts) を確認することができる。

STS ファイルは自動的に作成され、あらかじめファイル名を指定することはできない。

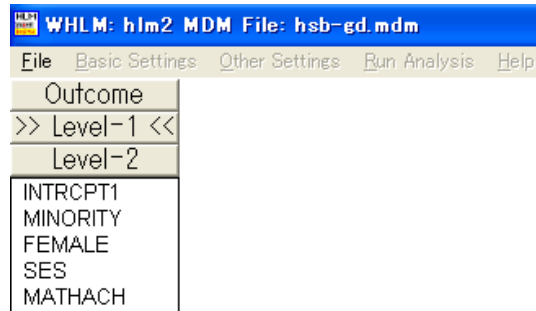
LEVEL-1 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
MINORITY	7185	0.27	0.45	0.00	1.00
FEMALE	7185	0.53	0.50	0.00	1.00
SES	7185	0.00	0.78	-3.76	2.69
MATHACH	7185	12.75	6.88	-2.83	24.99

LEVEL-2 DESCRIPTIVE STATISTICS					
VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
SIZE	160	1097.83	629.51	100.00	2713.00
SECTOR	160	0.44	0.50	0.00	1.00
PRACAD	160	0.51	0.26	0.00	1.00
DISCLIM	160	-0.02	0.98	-2.42	2.76
HIMINTY	160	0.28	0.45	0.00	1.00
MEANSES	160	-0.00	0.41	-1.19	0.83

3.5 モデルの指定

Done

[Done] ボタンをクリックすることで MDM ファイルの作成段階は終了し、モデルの指定を行う画面が現れる。



3.5 モデルの指定

MDM ファイルを作成したあとは、それを元に実際に分析したい統計モデルを指定することになる。

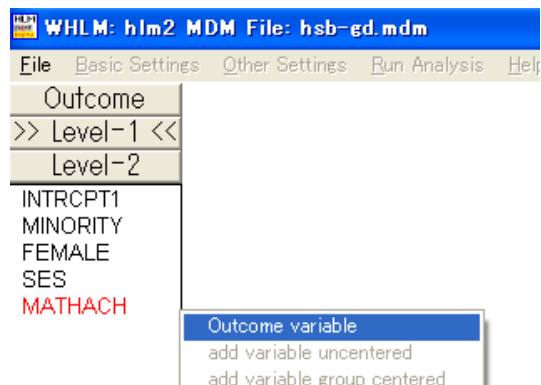
HLM6 では、式 (1) から式 (3) に従って変数を指定していけばモデルの指定ができる。こうした直感的な操作を可能にする優れた GUI 機能を持つことが HLM6 の大きな特徴となっている。

以前に作成した MDM ファイルを再利用する際は、[File]-[Create a new model using an existing MDM file] メニューから MDM ファイルを選択すれば、モデル指定の画面に移る。

3.5.1 レベル 1 のモデル式

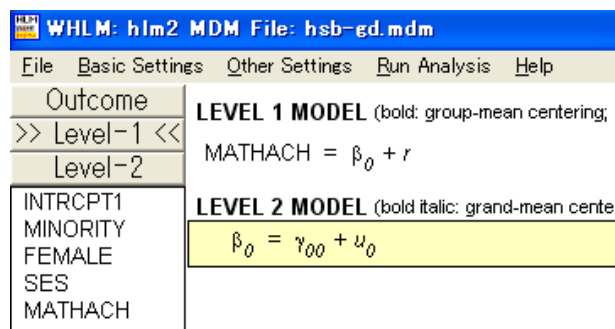
式 (1) のとおり、ここでは、生徒の数学の成績 (MATHACH) をその生徒の家庭の社会的経済的地位 (SES) で説明するモデル式を立てる。

まず、[Level-1] ボタンをクリックして、レベル 1 の入力モードにする (通常はすでにになっている)。下にレベル 1 のモデル式に投入するための変数の一覧が表示されているので、そこから "MATHACH" をクリックし、"Outcome variable" を選択する。

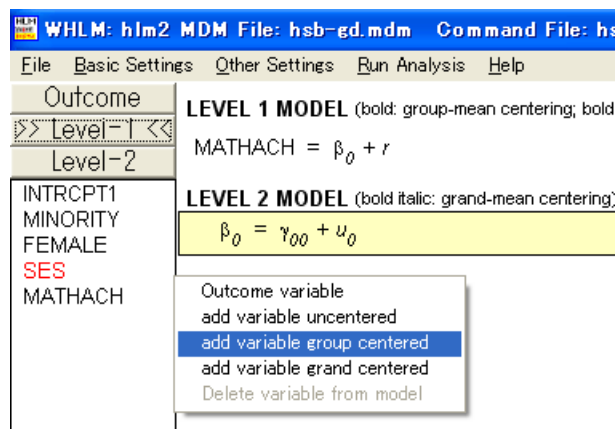


これで、"MATHACH" が従属変数に入ったモデル式が表示される。

3.5 モデルの指定

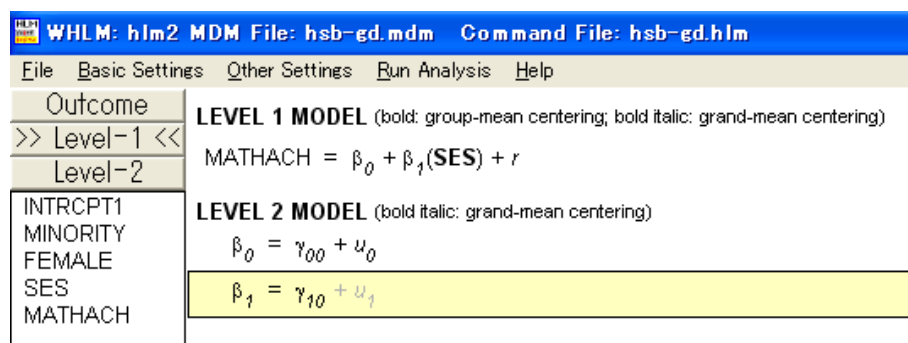


次に，“SES”をクリックして，“add variable group centered”を選択する。これを選択した場合，説明変数の部分が太字（ボールド体）で表示される。



この場合，切片 β_0 は各学校における数学成績の平均の推定値となる。これは，レベル 2 モデルに SES の平均である“MEANSES”を入れることから，多重共線性を避けるための措置でもある。HLM6 では，その他にも全体平均でのセンタリングを行うこともできる。センタリングを行った際は，切片の意味が代わってくることに注意する必要がある。

レベル 1 のモデル式に説明変数として“SES”が投入されたのが確認できる。もし間違っただけの変数を投入してしまった際は，その変数をもう一度選択し，“Delete variable from model”を選択すればよい。



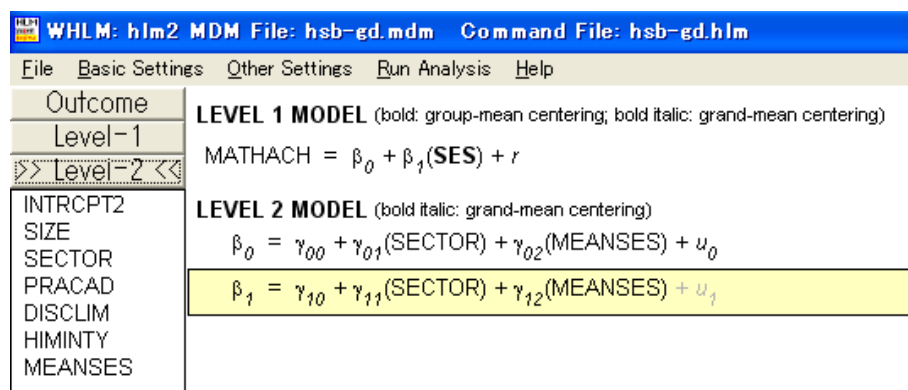
レベル 2 のモデル式

[Level-2] をクリックして，レベル 2 のモデルの入力モードに切り替える。切り替えた直後は，レベル 1 のモデル式の切片（ β_0 ）を説明するモデル式の部分が黄色くなっており，まずこの部分の入力を行う。式 (2) を参照し，先ほどと同じよ

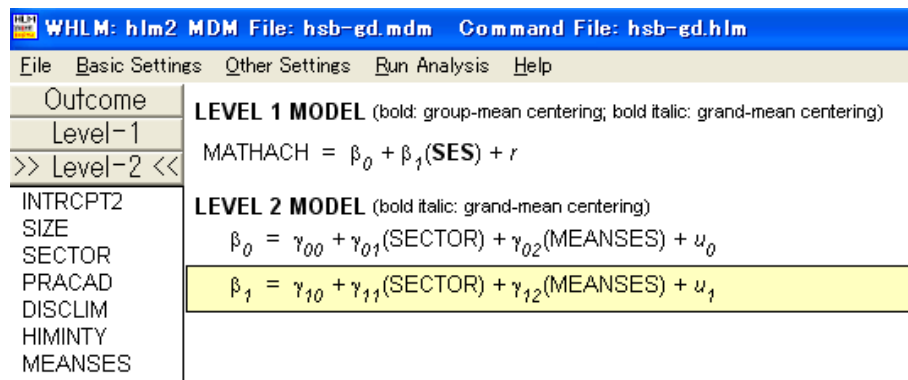
3.6 オプションの指定

うに”SECTOR”および”MEANSES”を説明変数として投入する（ここではともに”uncenterd”を選択。この場合説明変数は細字で表示される。）

次に、 β_1 を説明するモデル式のエリアをクリックして、ここを黄色くさせる。式(3)を参照し、先ほどと同じように”SECTOR”および”MEANSES”を説明変数として投入する（ここではともに”uncenterd”を選択。）



ここで、よく見るとレベル2の残差(u_1)が灰色になっているのに気づく。これは残差成分がモデル式に投入されていないことを示しているので、この部分をクリックして黒くする。これで u_1 がモデルに投入された。



同様に、レベル2の他の誤差成分についてもその部分をクリックすることでモデルからはずしたり投入しなおしたりすることができる。レベル1の誤差成分ははずすことはできない。

3.6 オプションの指定

次に、さまざまなオプションの設定を行う。

3.6.1 Basic Settings

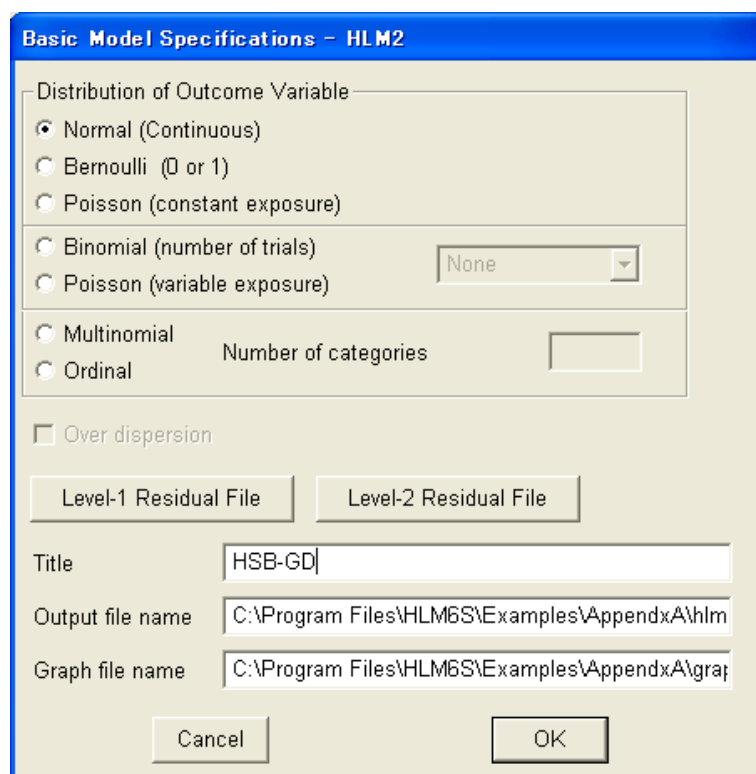
[Basic Settings] メニューでは、従属変数の分布に関する指定、結果のファイルの名前、分析のタイトル、残差ファイル指定などを行うことができる。

まず、[Basic Setting] をクリックする。

オプションであるので、必須というわけではないが、残差ファイルに関してはここで指定しないと作成されない。

もしくは、[Level-1] の上の [Outcome] をクリックしてもよい。

3.6 オプションの指定

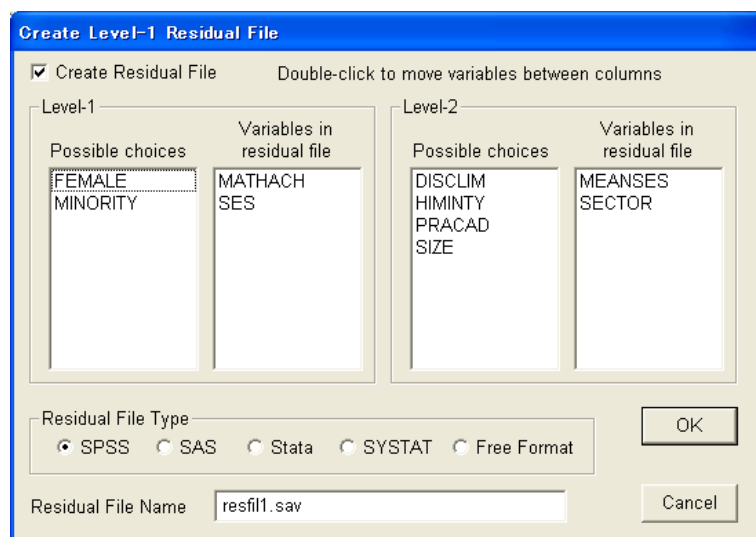


The dialog box is titled "Basic Model Specifications - HLM2". It contains several sections for specifying model parameters. The "Distribution of Outcome Variable" section has radio buttons for "Normal (Continuous)", "Bernoulli (0 or 1)", "Poisson (constant exposure)", "Binomial (number of trials)", "Poisson (variable exposure)", "Multinomial", and "Ordinal". The "Binomial" and "Poisson (variable exposure)" options have a dropdown menu set to "None". The "Multinomial" option has a text box for "Number of categories". There is a checkbox for "Over dispersion". Below these are buttons for "Level-1 Residual File" and "Level-2 Residual File". The "Title" field contains "HSB-GD". The "Output file name" field contains "C:\Program Files\HLM6S\Examples\Appendx\hlm". The "Graph file name" field contains "C:\Program Files\HLM6S\Examples\Appendx\A\graf". At the bottom are "Cancel" and "OK" buttons.

残差ファイルの指定

ここでは、残差を出力するファイルの指定について説明する。残差ファイルには、各レベルの残差だけでなくレベル1の切片や傾きの推定値など重要な情報がたくさん含まれている。残差ファイルの設定は、各レベルごとに行う。

まず、レベル1の残差ファイルの設定を行うために、[Level-1 Residual File] をクリックする。



The dialog box is titled "Create Level-1 Residual File". It has a checkbox "Create Residual File" which is checked. Below it is a section for "Level-1" with two columns: "Possible choices" and "Variables in residual file". The "Possible choices" column contains "FEMALE" and "MINORITY". The "Variables in residual file" column contains "MATHACH" and "SES". To the right is a section for "Level-2" with similar columns. The "Possible choices" column contains "DISCLIM", "HIMINTY", "PRACAD", and "SIZE". The "Variables in residual file" column contains "MEANSES" and "SECTOR". Below these sections is a "Residual File Type" section with radio buttons for "SPSS", "SAS", "Stata", "SYSTAT", and "Free Format". The "SPSS" option is selected. At the bottom is a "Residual File Name" field containing "resfil1.sav". There are "OK" and "Cancel" buttons.

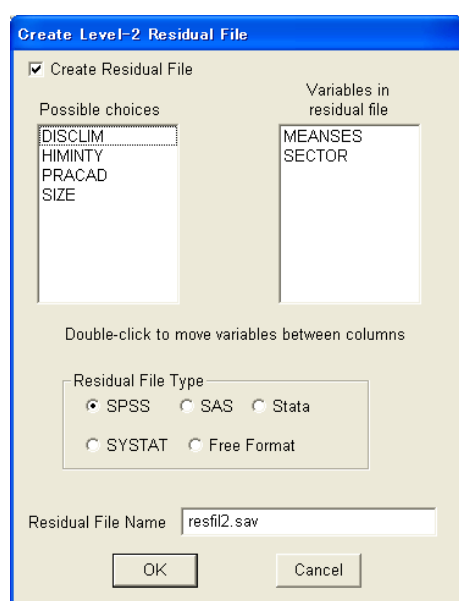
3.6 オプションの指定

レベル1，レベル2それぞれの”Variables in residual file”に，モデルの設定のところで投入した変数がそのまま入っているはずである。もし必要であればさらに変数をダブルクリックして投入することができる。ただし，すでにモデルに投入済みの変数は取り除くことができない。

ここではデフォルトのままとし，”Residual File Type”で”SPSS”を選択した。ファイル名もデフォルトのまま，”resfil1.sav”とした。

[OK] をクリックしてこのウィンドウを閉じる。

次に，レベル2 モデルの残差ファイルを指定する。[Level-2 Residual File] をクリックする。



ここで仮に”Free Format”を選択し，テキストファイル(.txt)やCSV ファイル(.csv)で保存すると，カンマ区切りの残差ファイルが作成される。SPSS がなければこのようにして作成した残差ファイルをあとで Excel などで開けばよい。(テキストエディタで開いてもよいが，見づらい。)

ここでも，”Variables in residual file”に，モデルの設定のところで投入したレベル2の変数がそのまま入っているはずである。レベル1 のときと同様，変数に関してはデフォルトのままとし，”Residual File Type”で”SPSS”を選択し，[OK] をクリックする。

これで，分析後”resfil1.sav”および”resfil2.sav”という2つの残差ファイルができるはずである。

タイトルと結果の出力ファイル名の指定

”Title”に，この分析のタイトルを入れる。

”Output file name”に，出力ファイルの名前をフルパスで入力する。ここでは，デフォルトのままとした。

”Graph file name”に，グラフを作る際のファイルの名前をフルパスで入力する。ここでは，デフォルトのままとした。

[OK] をクリックして，終了。

デフォルトでは”no title”となっている。このままでもよいが，今回は”HSB-GD”とした。

3.7 HLM ファイルの保存

3.6.2 Other Settings

ここでは、反復計算、推定法、統計的検定に関する設定などを行うことができる。
今回はデフォルトのままにしておく。

3.7 HLM ファイルの保存

ここまで設定した内容は、実際に分析を行う前に HLM ファイル (.hlm) として保存する必要がある。

[File]-[Save As] で適切なファイル名を指定し、HLM ファイルを保存する。

分析モデルのタイプとして HMLM を選択した場合は、拡張子を ".mlm" として保存するといよい。

ここでは、"hsb-gd.hlm" とした。

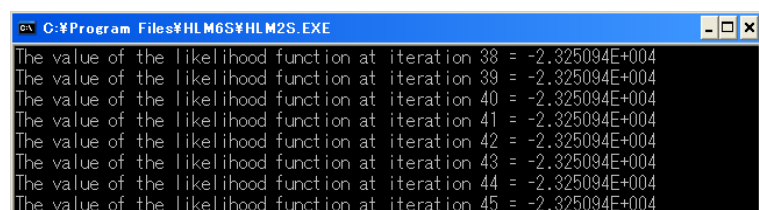
3.8 分析の実行

[Run Analysis] をクリックする。

反復計算の過程を表す画面が表示される。計算が終わると、自動的に消える。

もし HLM ファイルを保存し忘れた場合は、ここで保存するよう促される。

もし指定した回数よりも多くの反復計算を必要とした場合は、所定の回数計算し終わった後、さらに続けるかどうか確認される。

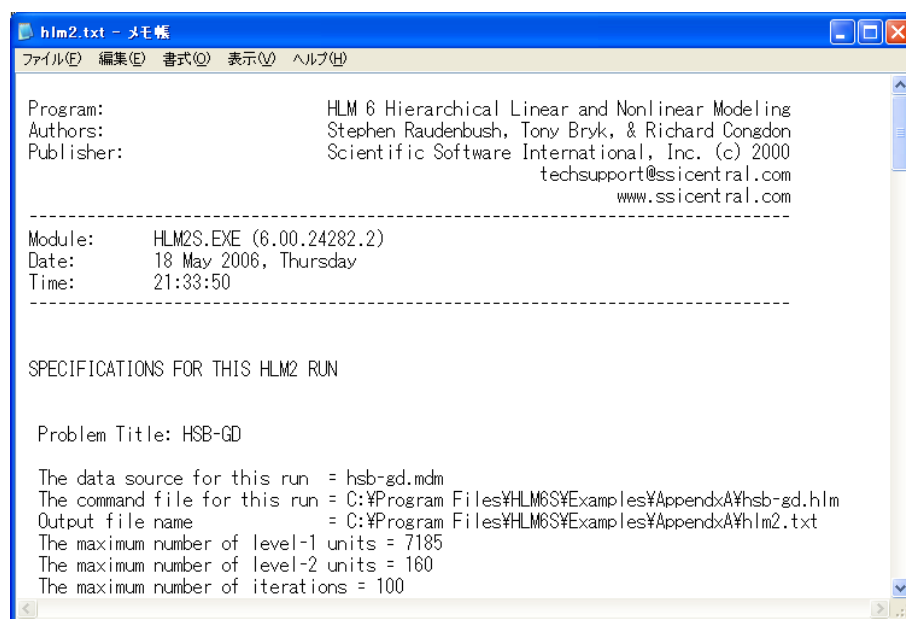


```
C:\Program Files\HLM6S\HLM2S.EXE
The value of the likelihood function at iteration 38 = -2.325094E+004
The value of the likelihood function at iteration 39 = -2.325094E+004
The value of the likelihood function at iteration 40 = -2.325094E+004
The value of the likelihood function at iteration 41 = -2.325094E+004
The value of the likelihood function at iteration 42 = -2.325094E+004
The value of the likelihood function at iteration 43 = -2.325094E+004
The value of the likelihood function at iteration 44 = -2.325094E+004
The value of the likelihood function at iteration 45 = -2.325094E+004
```

3.9 結果の確認

分析はここで終了したので、結果を確認する。[File]-[View Output] とクリックすると、結果を格納したテキストファイルが自動的に立ち上がる。

今回は [Basing Settings] において結果のファイル名をデフォルトのままとしたので、ここでは "hlm2.txt" となっている。もしファイル名を指定していれば、そのファイルが自動的に立ち上がる。



```
hlm2.txt - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)

Program:      HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors:      Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:     Scientific Software International, Inc. (c) 2000
               techsupport@ssicentral.com
               www.ssicentral.com

-----
Module:       HLM2S.EXE (6.00.24282.2)
Date:         18 May 2006, Thursday
Time:         21:33:50
-----

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: HSB-GD

The data source for this run = hsb-gd.mdm
The command file for this run = C:\Program Files\HLM6S\Examples\AppendxA\hsb-gd.hlm
Output file name = C:\Program Files\HLM6S\Examples\AppendxA\hlm2.txt
The maximum number of level-1 units = 7185
The maximum number of level-2 units = 160
The maximum number of iterations = 100
```

3.9 結果の確認

以下では、このファイルを見ながら結果の解釈を行うことになる。

以下, ”#” ~ ”#” で囲んだ部分は、
後から付け加えたコメントである。

基本的な情報

```
SPECIFICATIONS FOR THIS HLM2 RUN

# 分析のタイトル #
  Problem Title: HSB-GD

# 分析に使用した MDM ファイル #
  The data source for this run  = hsb-gd.mdm

# モデルの指定を行った .hlm ファイル #
  The command file for this run = C:\Program Files\HLM6S\Examples\AppendxA\hsb-gd.hlm

# 出力ファイルの名前 #
  Output file name              = C:\Program Files\HLM6S\Examples\AppendxA\hlm2.txt

# サンプルサイズおよび指定した反復回数 #
  The maximum number of level-1 units = 7185
  The maximum number of level-2 units = 160
  The maximum number of iterations = 100

# 分散・共分散行列の推定の方法（デフォルトでは、制約つき最尤推定。） #
  Method of estimation: restricted maximum likelihood

# 重み付けに関する情報 #
Weighting Specification
-----
                                Weight
                                Variable
                                Name      Normalized?
Level 1          no
Level 2          no
Precision        no
```

モデルの確認

```
# 従属変数は MATHACH #
  The outcome variable is  MATHACH

# モデルに含まれる切片および傾き（偏回帰係数）をレベルごとに表示 #
  The model specified for the fixed effects was:
-----

Level-1                      Level-2
Coefficients                  Predictors
-----
      INTRCPT1, B0          INTRCPT2, G00
                              SECTOR, G01
                              MEANSES, G02
*      SES slope, B1        INTRCPT2, G10
                              SECTOR, G11
                              MEANSES, G12
```

3.9 結果の確認

```
# SES は集団平均でセンタリングした。 #
'*' - This level-1 predictor has been centered around its group mean.

# 分散, 共分散行列に関する指定 #
The model specified for the covariance components was:
-----

Sigma squared (constant across level-2 units)

Tau dimensions
  INTRCPT1
    SES slope

# 指定したモデル式を提示 #
Summary of the model specified (in equation format)
-----

Level-1 Model

Y = B0 + B1*(SES) + R

Level-2 Model
B0 = G00 + G01*(SECTOR) + G02*(MEANSES) + U0
B1 = G10 + G11*(SECTOR) + G12*(MEANSES) + U1
```

分散・共分散行列の推定結果

```
# 収束するまでの反復数 #
Iterations stopped due to small change in likelihood function
***** ITERATION 61 *****

# レベル 1 の分散の推定値 #
Sigma_squared =      36.70313

# レベル 2 の分散・共分散行列の推定値 #
Tau
INTRCPT1,B0      2.37996      0.19058
    SES,B1      0.19058      0.14892

# レベル 2 の分散・共分散行列を相関行列に直したもの #
Tau (as correlations)
INTRCPT1,B0  1.000  0.320
    SES,B1  0.320  1.000

# レベル 1 の切片と傾きの推定値の信頼性 (Reliability) #
-----
Random level-1 coefficient  Reliability estimate
-----
INTRCPT1, B0                0.733
    SES, B1                0.073
-----
```

信頼性の値が 1 に近いほど、レベル 1 の切片と傾きを推定する際にサンプル全体の情報に対して当該集団の情報を利用する割合が大きくなる。信頼性は集団ごとに算出されるものであるが、ここでは全集団にわたって平均した値が報告されている。#

3.9 結果の確認

レベル 2 の切片と偏回帰係数の推定値

最終的な尤度関数の値

The value of the likelihood function at iteration 61 = -2.325094E+004

レベル 2 の切片と偏回帰係数の推定値と検定統計量, 有意性検定の結果

The outcome variable is MATHACH

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.096006	0.198734	60.865	157	0.000
SECTOR, G01	1.226384	0.306272	4.004	157	0.000
MEANSES, G02	5.333056	0.369161	14.446	157	0.000
For SES slope, B1					
INTRCPT2, G10	2.937981	0.157135	18.697	157	0.000
SECTOR, G11	-1.640954	0.242905	-6.756	157	0.000
MEANSES, G12	1.034427	0.302566	3.419	157	0.001

The outcome variable is MATHACH

最高次のレベルのサンプルサイズが極めて大きい場合は, (特に分散共分散行列に関する設定が間違っているかもしれないことを見越して) こちらの結果を採用するのが望ましいとされている。今回はこちらを見ればよい。(こちらを参照する必要がない場合はそのような指示が表示される。)#

Final estimation of fixed effects

(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.096006	0.173699	69.638	157	0.000
SECTOR, G01	1.226384	0.308484	3.976	157	0.000
MEANSES, G02	5.333056	0.334600	15.939	157	0.000
For SES slope, B1					
INTRCPT2, G10	2.937981	0.147620	19.902	157	0.000
SECTOR, G11	-1.640954	0.237401	-6.912	157	0.000
MEANSES, G12	1.034427	0.332785	3.108	157	0.003

最終的な分散の推定値

最終的な分散の推定値と検定結果

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	1.54271	2.37996	157	605.29503	0.000
SES slope, U1	0.38590	0.14892	157	162.30867	0.369
level-1, R	6.05831	36.70313			

3.10 グラフの表示など

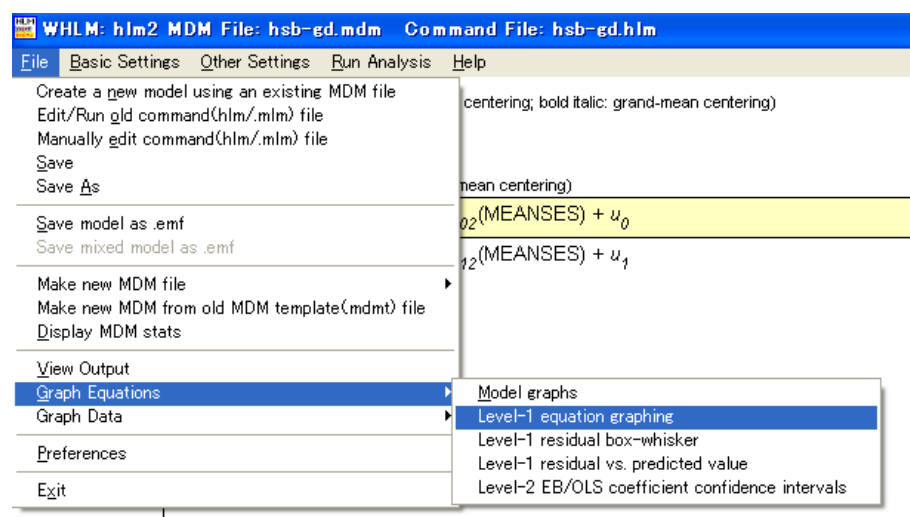
Deviance など

```
# Deviance #
Statistics for current covariance components model
-----
Deviance                      = 46501.875643
# 推定した分散・共分散の数 #
Number of estimated parameters = 4
```

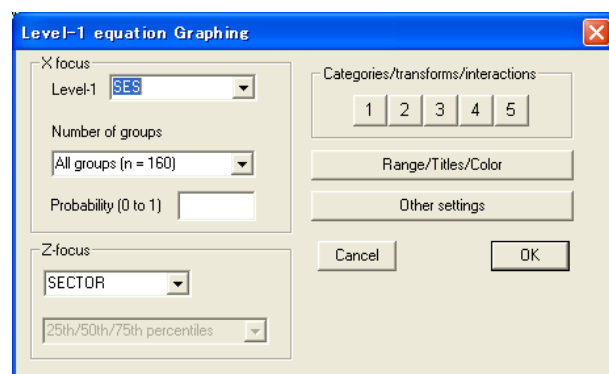
3.10 グラフの表示など

HLM6 には、さまざまなグラフを表示させる機能がついている。ここでは、例として式 (1) で示したレベル 1 の回帰式が、学校の属性を表す SECTOR によってどのように異なるのかグラフで示してみたい。

[File]-[Graph Equations]-[Level-1 equation graphing] を選択する。



レベル 1 のモデル式をグラフ化するためのウィンドウが立ち上がる。



3.10 グラフの表示など

X-focus

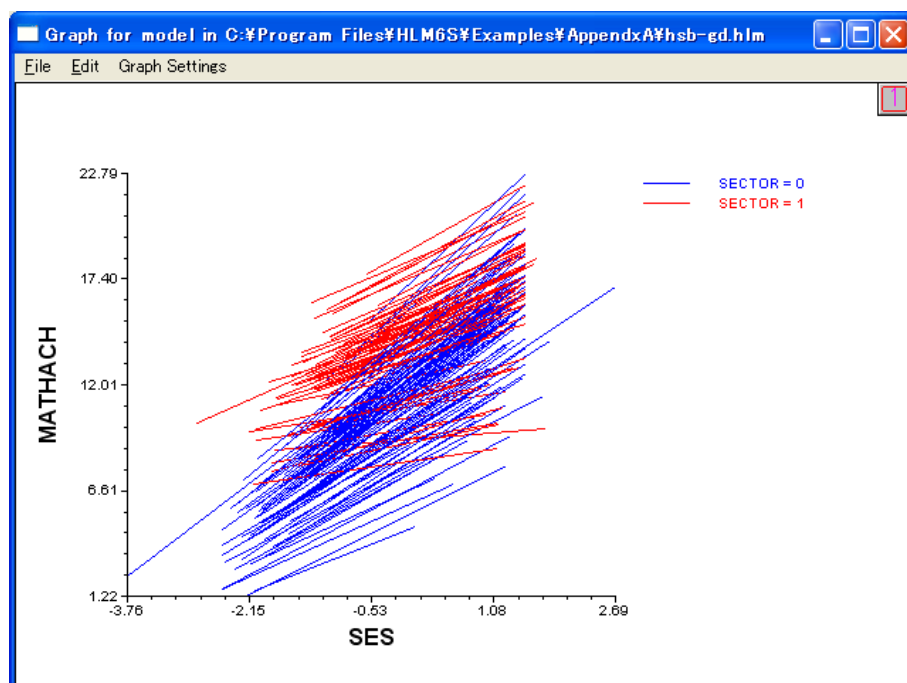
Level-1: レベル 1 の説明変数を投入する。ここでは, "SES" を選択。

Number of groups: 表示するグループの数を選択する。ここでは, "All groups" を選択。

Z-focus

学校を分類する変数を選択。ここでは, "SECTOR" を選択。

[OK] をクリックすると, グラフが表示される。



このグラフを見ると, Cathoric (SECTOR=1) の学校のほうが切片の値は大きい傾きは小さい傾向にあることがわかる。これは SECTOR にかかるレベル 2 の偏回帰係数の推定値の符号と一致している。

ここで, 21 ページに示したレベル 2 の偏回帰係数の推定結果の欄を次に再掲したのでグラフとの対応を確認されたい。

変数 SECTOR にかかる偏回帰係数の値に関して, 切片 ("INTRCPT1, B0") 欄では 1.226384 とプラスの値であるのに対し, 傾き ("SES slope, B1") 欄では -1.640954 とマイナスの値となっていることがわかる。

3.11 残差のチェック

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.096006	0.173699	69.638	157	0.000
SECTOR, G01	1.226384	0.308484	3.976	157	0.000
MEANSES, G02	5.333056	0.334600	15.939	157	0.000
For SES slope, B1					
INTRCPT2, G10	2.937981	0.147620	19.902	157	0.000
SECTOR, G11	-1.640954	0.237401	-6.912	157	0.000
MEANSES, G12	1.034427	0.332785	3.108	157	0.003


3.11 残差のチェック

HLM6 では、結果の出力にレベル 1 の（この場合は学校ごとの）切片や傾きの推定値は出力されない。これらについては残差ファイルの中に格納されているため、必要であればこちらを確認する必要がある。また、残差の確認はモデルの仮定が成り立っているかどうかをチェックするためにも有効である。

デフォルトでは、MDM ファイルを格納したフォルダに残差ファイルが格納されているはずである。

もちろん分析の際に指定しなければ作成されない。

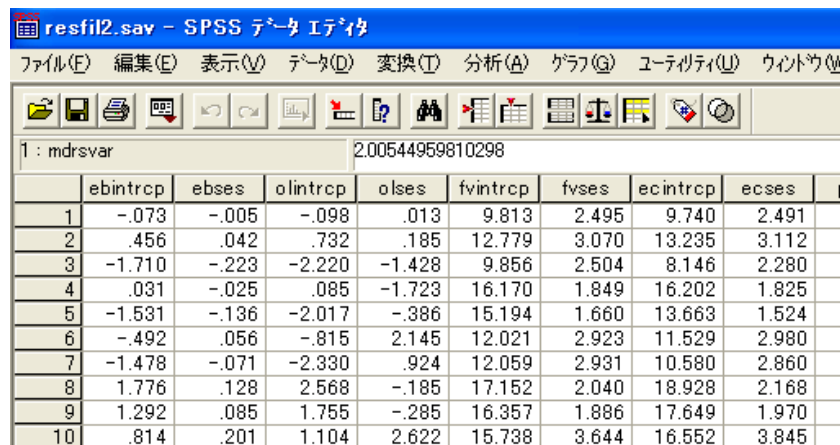
3.11.1 レベル 1 の残差ファイル



	l2id	l1resid	fitval	sigma	ses	mathach	sector	meanses	va
1	1224	-1.140	7.016	6.058	-1.094	5.876	.000	-.428	
2	1224	10.351	9.357	6.058	-.154	19.708	.000	-.428	
3	1224	10.842	9.507	6.058	-.094	20.349	.000	-.428	
4	1224	-.377	9.158	6.058	-.234	8.781	.000	-.428	
5	1224	7.470	10.428	6.058	.276	17.898	.000	-.428	
6	1224	-6.294	10.877	6.058	.456	4.583	.000	-.428	
7	1224	-12.115	9.283	6.058	-.184	-2.832	.000	-.428	
8	1224	-7.813	8.336	6.058	-.564	.523	.000	-.428	
9	1224	-7.083	8.610	6.058	-.454	1.527	.000	-.428	
10	1224	11.840	9.681	6.058	-.024	21.521	.000	-.428	

”l1resid”とあるのが、レベル 1 の残差である。”fitval”がモデル式によって予測された従属変数の値であり、これら 2 つの値を合計すると”mathach”と等しくなる。

3.11.2 レベル 2 の残差ファイル



	ebintcrp	ebses	olintcrp	olse	fvintcrp	fvs	ecintcrp	eces	f
1	-.073	-.005	-.098	.013	9.813	2.495	9.740	2.491	
2	.456	.042	.732	.185	12.779	3.070	13.235	3.112	
3	-1.710	-.223	-2.220	-1.428	9.856	2.504	8.146	2.280	
4	.031	-.025	.085	-1.723	16.170	1.849	16.202	1.825	
5	-1.531	-.136	-2.017	-.386	15.194	1.660	13.663	1.524	
6	-.492	.056	-.815	2.145	12.021	2.923	11.529	2.980	
7	-1.478	-.071	-2.330	.924	12.059	2.931	10.580	2.860	
8	1.776	.128	2.568	-.185	17.152	2.040	18.928	2.168	
9	1.292	.085	1.755	-.285	16.357	1.886	17.649	1.970	
10	.814	.201	1.104	2.622	15.738	3.644	16.552	3.845	

レベル 2 の残差ファイルにはかなり多くの情報が格納されているので、重要と思われるものを以下にピックアップして説明した。

”eb・・・”： ”ebintcrp” とあるのは、レベル 1 の切片 (β_{0j}) の推定に経験ベイズ (empirical Bayes) 法を用いた際の残差 (u_{0j}) である。同様に ”ebses” とあるのは、レベル 1 の傾き (β_{1j}) に経験ベイズ法を用いた際の残差 (u_{1j}) である。

”ol・・・”： ”olintcrp”, ”olse” とあるのは、それぞれレベル 1 の切片 (β_{0j}) および傾き (β_{1j}) を (学校ごとに) 通常の最小 2 乗 (ordinary least-squares) 法で推定した際の残差 (u_{0j} および u_{1j}) である。この方法では経験ベイズ法の場合と異なりサンプル全体の情報を使っていないので、”eb・・・”で示した残差よりもこちらの方が残差 (の絶対値) は大きくなっている。

”fv・・・”： それぞれレベル 1 の切片と傾きのレベル 2 モデルによる予測値である。

”ec・・・”： それぞれレベル 1 の切片と傾きの経験ベイズ推定値である。これは、予測値 (”fv・・・”) に残差の値 (”eb・・・”) を足したものに等しい。

経験ベイズ法では、レベル 1 の切片と傾きについて、学校ごとに最小 2 乗法を用いて推定した値 (次に述べる ”ol・・・” で示された値) とレベル 2 のモデル式にもとづく推定値とをそれぞれの信頼性で重み付けした値を最終的な推定値とする。

”ebintcrp” と ”olintcrp”, ”ebses” と ”olse” の値を実際に比較されたい。

今回の分析例では、学校ごとの切片 (β_{0j}) および傾き (β_{1j}) の推定値となる。23 ページで示したグラフは、これらの値を用いて描かれたものである。

4 SAS PROC MIXED による分析

階層的線形モデルは、何も HLM6 を用いてしか分析できないというわけではない。基本的に階層的線形モデルは線形混合モデルと考えることができるので、線形混合モデルを扱うソフトウェアであれば階層的線形モデルによる分析を行うことができる。

ここでは、SAS において混合モデルを扱うための MIXED プロシジャを用いて階層的線形モデルによる分析を行う方法について簡単に説明する。

SAS PROC MIXED では、HLM6 とモデル式の表し方が若干異なる。

ただし、検定の方法や反復計算の手法などが多少異なる場合がある。また、HLM6 と異なり、この場合はデータセットを 1 つのファイルにまとめる必要がある。(具体的なデータは、29 ページの R で用いたものと同様のものであるので参照されたい。) 以下の例は、すでに整形したデータファイルを使用して分析したものである。

4.1 SAS プログラム

式 (2) および式 (3) を式 (1) に代入して整理すると,

$$\begin{aligned} MATHACH_{ij} = & [\gamma_{00} + \gamma_{01}SECTOR_j + \\ & \gamma_{02}MEANSES_j + \gamma_{10}(SES_{ij} - \bar{SES}_{.j}) + \\ & \gamma_{11}SECTOR_j(SSES_{ij} - \bar{SES}_{.j}) + \\ & \gamma_{12}MEANSES_j(SSES_{ij} - \bar{SES}_{.j})] + \\ & [u_{0j} + u_{1j}(SES_{ij} - \bar{SES}_{.j}) + r_{ij}] \end{aligned} \quad (4)$$

となる。このうち、最初の [] 内は固定効果であり、2 番目の [] 内は変量効果である。HLM6 ではレベルごとに分けた式を入力できたのに対し、SAS (およびこの後で触れる R) においては、すべてのレベルを統合したモデル式 (4) をもとに固定効果と変量効果の部分に分けて記述することになる。SAS PROC MIXED を用いた階層的モデルの分析方法について、詳しくは、Singer (1998) を参照されたい。

以下に、HSB データを用いて HLM6 を用いて行ったのと同じ分析をするためのプログラムの一例とその結果の出力を示した。

4.1 SAS プログラム

```
*****;
* SAS による HSB データの分析 ;
*   By Taichi OKUMURA   ;
*****;

OPTIONS ls = 80;

/* データの読み込み */
DATA hsb;
INFILE 'C:\Program Files\SAS Institute\SAS\V8\hsb-sas.dat';
INPUT id school ses mathach meanses cses sector;
RUN;

/* proc mixed による分析 */
PROC MIXED DATA=hsb NOCLPRINT COVTEST;
CLASS school;
MODEL mathach = sector
          meanses
          cses
          sector*cses
          meanses*cses / SOLUTION DDFM=bw NOTEST;
RANDOM intercept cses / TYPE = un SUB=school;
RUN;
QUIT;
```

以下に、MIXED プロシジャの部分を簡単に説明しておく。

PROC MIXED

NOCLPRINT: 集団 (この場合は school) ごとの結果は出力しない

COVTEST: 分散要素の検定を行う

4.2 SAS のアウトプット

CLASS

school は数値ではなく、カテゴリ変数。

MODEL

この部分には固定効果の式を書く。

SOLUTION: 固定効果の推定値を出力させる

DDFM=bw: 固定効果の検定の際の自由度の設定

NOTEST: 固定効果に関する仮説検定を行わない

RANDOM

この部分には変量効果の式を書く。

TYPE = un: 分散・共分散行列に構造を仮定しない

sub=school: 今回の例では、個々の生徒の観測値は学校（"school"）にネストされている

4.2 SAS のアウトプット

上記のプログラムを実行したときのアウトプットを以下に示す。

"/*" ~ "*/"の部分はコメントである。

SAS システム

The Mixed Procedure

Model Information

Data Set	WORK.HSB
Dependent Variable	mathach
Covariance Structure	Unstructured
Subject Effect	school
Estimation Method	REML /* REML 推定 (デフォルト) */
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Dimensions

Covariance Parameters	4
Columns in X	6
Columns in Z Per Subject	2
Subjects	160
Max Obs Per Subject	67
Observations Used	7185
Observations Not Used	0
Total Observations	7185

4.2 SAS のアウトプット

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	46724.22627510	
1	2	46503.66454957	0.00000010
2	1	46503.66286827	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
/* レベル 2 の分散・共分散の推定値と有意性検定の結果 */					
UN(1,1)	school	2.3794	0.3714	6.41	<.0001
UN(2,1)	school	0.1918	0.2043	0.94	0.3479
UN(2,2)	school	0.1012	0.2138	0.47	0.3180
/* レベル 1 の分散の推定値と有意性検定の結果 */					
Residual		36.7212	0.6261	58.65	<.0001

Fit Statistics

-2 Res Log Likelihood	46503.7	/* deviance */
AIC (smaller is better)	46511.7	
AICC (smaller is better)	46511.7	
BIC (smaller is better)	46524.0	

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	220.56	<.0001

/* 固定効果 (レベル 2 の偏回帰係数) の推定結果・有意性検定 */

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t	
Intercept	12.1279	0.1993	157	60.86	<.0001	/* G00 */
sector	1.2266	0.3063	157	4.00	<.0001	/* G01 */
meanses	5.3329	0.3692	157	14.45	<.0001	/* G02 */
cses	2.9450	0.1556	7022	18.93	<.0001	/* G10 */
sector*cses	-1.6427	0.2398	7022	-6.85	<.0001	/* G11 */
meanses*cses	1.0392	0.2989	7022	3.48	0.0005	/* G12 */

5 R による分析

フリーの統計解析ソフトである R (R Development Core Team, 2005) には、混合モデルを分析するためのパッケージ `nlme` が用意されている。このパッケージにある `lme` 関数を用いることで、階層的線形モデルと同様の分析を行うことができる^{*5}。

`nlme` パッケージをインストールすると、サンプルデータとして HSB データがついてくるので、それを用いて今回行ったのと同じ分析を行うことができる。

ただし、以下のプログラムで示したようにデータを整形する必要がある。

5.1 データ

SAS の場合と同様、データファイルは 1 つにまとめる必要がある。以下に、HSB のデータを整形したものの最初の 10 ケース目までを示した。

	school	ses	mathach	meanses	cses	sector
1	1224	-1.528	5.876	-0.434383	-1.09361702	Public
2	1224	-0.588	19.708	-0.434383	-0.15361702	Public
3	1224	-0.528	20.349	-0.434383	-0.09361702	Public
4	1224	-0.668	8.781	-0.434383	-0.23361702	Public
5	1224	-0.158	17.898	-0.434383	0.27638298	Public
6	1224	0.022	4.583	-0.434383	0.45638298	Public
7	1224	-0.618	-2.832	-0.434383	-0.18361702	Public
8	1224	-0.998	0.523	-0.434383	-0.56361702	Public
9	1224	-0.888	1.527	-0.434383	-0.45361702	Public
10	1224	-0.458	21.521	-0.434383	-0.02361702	Public

5.2 R プログラム

```
#-----#
#                               R による HSB データの分析
#                               By Taichi OKUMURA
#-----#

# nlme パッケージの読み込み
library(nlme)

# レベル 1 のデータの読み込み
data(MathAchieve)
MathAchieve[1:10,]

# レベル 2 のデータの読み込み
data(MathAchSchool)
MathAchSchool[1:10,]

# 学校ごとの SES の平均を msas という新しい変数にする。
attach(MathAchieve)
```

^{*5} <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion/appendix-mixed-models.pdf> に詳しい解説がある。

5.3 R のアウトプット

```
mres <- tapply(SSES, School, mean)
detach(MathAchieve)

# HSB というデータセット
HSB <- as.data.frame(MathAchieve[, c("School", "SES", "MathAch")])
names(HSB) <- c("school", "ses", "mathach")

# 学校ごとの SES の平均を mses という新しい変数にする。
HSB$meanses <- mres[as.character(HSB$school)]
HSB$cres <- HSB$ses - HSB$meanses

# sector 変数を取り出して HSB に投入
sector <- MathAchSchool$Sector
names(sector) <- row.names(MathAchSchool)
HSB$sector <- sector[as.character(HSB$school)]
HSB$sector <- factor(HSB$sector, levels=c('Public', 'Catholic'))

# result.hsb に分析結果を格納
result.hsb <- lme(mathach ~ sector + meanses + cres +
                  sector*cres + meanses*cres,
                  random = ~ cres | school,
                  data = HSB)

# 結果の表示
summary(result.hsb)
```

5.3 R のアウトプット

上記のプログラムを実行したときのアウトプットを以下に示す。

”/*” ~ ”*/”の部分はコメントである。

```
Linear mixed-effects model fit by REML /* REML 推定 (デフォルト) */
Data: HSB
      AIC      BIC    logLik
46523.66 46592.45 -23251.83 /* LogLik (対数尤度) の値に-2 をかけたものが deviance */

Random effects:
Formula: ~cres | school
Structure: General positive-definite, Log-Cholesky parametrization
          StdDev   Corr
(Intercept) 1.5426150 (Intr) /* レベル 2 の分散の平方根 (切片) */
cres         0.3182015 0.391 /* レベル 2 の分散の平方根 (傾き) */
Residual    6.0597955      /* レベル 1 の分散の平方根 */

/* レベル 2 の偏回帰係数の推定結果 */
Fixed effects: mathach ~ sector + meanses + cres + sector * cres + meanses * cres
              Value Std.Error   DF  t-value p-value
(Intercept)  12.127931 0.1992920 7022  60.85510  0e+00 /* G00 */
sectorCatholic  1.226579 0.3062733  157   4.00485  1e-04 /* G01 */
meanses        5.332875 0.3691684  157  14.44564  0e+00 /* G02 */
cres           2.945041 0.1556005 7022  18.92694  0e+00 /* G10 */
sectorCatholic:cres -1.642674 0.2397800 7022 -6.85076  0e+00 /* G11 */
meanses:cres     1.039230 0.2988971 7022   3.47688  5e-04 /* G12 */

Correlation:
          (Intr) sctrCt meanss cres   sctrC:
sectorCatholic -0.699
meanses        0.256 -0.356
cres           0.075 -0.053  0.019
```

REFERENCES

```
sectorCatholic:cse -0.052  0.077 -0.027 -0.696
meanses:cse        0.019 -0.026  0.074  0.293 -0.351

Standardized Within-Group Residuals:
      Min      Q1      Med      Q3      Max
-3.1592608 -0.7231893  0.0170471  0.7544510  2.9582205

Number of Observations: 7185 /* データに含まれる生徒の数 */
Number of Groups: 160      /* 学校の数 */
```

References

- R Development Core Team. (2005). *R: A language and environment for statistical computing*. Vienna, Austria. (ISBN 3-900051-07-0)
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 24, 323-355.