This implements /
REPEATED and /
RANDOM in SPSS
using Singer
(1998) as an
example. Excellent.

Designing Multilevel Models Using SPSS 11.5 Mixed Model

John Painter, Ph.D.

Jordan Institute for Families
School of Social Work
University of North Carolina at Chapel Hill

Creating Multilevel Models Using SPSS 11.5 Mixed Model

The purpose of this paper is to illustrate applications of SPSS Mixed Models syntax to multilevel modeling by replicating analyses presented by Singer (1998) using SAS Proc Mixed. It is not intended as a primer for understanding multilevel modeling per se. Readers interested in learning more about the statistical and conceptual basis of multilevel modeling should refer to Singer (1998) article for a complete discussion of the examples presented here, as well as books by Snijders & Bosker (2000) and Singer & Willett (2003). The advantage of using programs such as SPSS (and SAS) in multilevel modeling is the ability to efficiently manage and manipulate data within a single software program. This capability offers a significant saving of time while reducing potential for human error.

This paper is divided into two sections. The first section offers a brief overview of data file structure as it relates to multilevel modeling and a brief comparison of SPSS Mixed and SAS Proc Mixed syntax. The second section presents examples of growth modeling. In order to minimize space and maintain continuity illustration of how specific SPSS syntax can be generated via menu options is presented separately in the Appendix.

Data Structure and Syntax

<u>Data Structure</u>

Appropriate structure of the data file is an important yet often unmentioned condition in multilevel analysis. There are two basic structures to consider in multilevel modeling: the “Multiple-Variable” (MV), and ‘Multiple-Record” (MR) (Singer, 1998).

The defining characteristic of Multiple-Variable (MV) structure is that all information pertaining to a single observation is placed on one line in the dataset. For example, if there are 20 participants in a study with 3 variables recorded for each person, the resulting MV data file will contain at least 20 lines (records) and 3 variables as illustrated in Figure 1. When multilevel modeling is employed it is also necessary to have at least one ‘nesting’ variable that identifies inherent group membership of multiple records. It is also common to include an individual level id as well. Multiple-Variable structure data is typically used with the SPSS Mixed model multilevel analysis.

Illustration 1. Multiple Variable Data Structure

ID	**Var1**	**Var2**	**Var3**	**Group**
1	12	45	34	A
2	23	43	34	B
3	31	54	45	C
4	13	42	31	A
5	26	40	38	B
6	27	49	44	C

The defining characteristic of Multiple-Record (MR) structure is that information pertaining to a single observation is ‘stacked’ or placed on multiple lines in the dataset. The MR structure is typical of longitudinal or repeated measures data. For example if there are 20 participants in a study with 3 observations on variable X for each person, the resulting MR data file will contain 120 lines (records) and 3 variables as illustrated in illustration 2. Note that in addition to variable X, it also <u>necessary</u> to include an individual level identifier (ID) and a variable representing timing or sequence (ORDER) of measurements.

Illustration 2. Multiple Record Data Structure

ID	**Var X**	**Order**
1	12	1
1	45	2
1	34	3
2	23	1
2	43	2
2	34	3
3	31	1
3	54	2
3	45	3

<u>Comparison of SPSS Mixed and SAS Proc Mixed syntax</u>
Table 1 provides code for the same equation expressed in both SPSS and SAS. Commands are typed in upper case, variables in lower case.

Table 1. Example of SPSS and SAS syntax

SPSS Mixed	GET FILE = ‘C:/hsb12.sav’ . MIXED mathach BY sector WITH meanses cses /METHOD = REML /PRINT = SOLUTION TESTCOV /FIXED = meanses sector cses meanses*cses sector*cses | SSTYPE(3) /RANDOM = INTERCEPT cses | SUBJECT(school) COVTYPE(UN).
SAS Proc Mixed	PROC MIXED DATA = hlmc.hsb12 COVTEST NOCLPRINT ; CLASS school sector ; MODEL mathach = meanses sector cses meanses*cses sector*cses / SOLUTION ; RANDOM INTERCEPT cses / SUB=school ; RUN;

As mentioned, a particularly useful aspect of SPSS Windows is the ease with which syntax commands can be produced by using menu options. In order to conserve space, two examples of creating code are presented in the Appendix. For a complete listing of the options available via the SPSS Mixed procedure see the Mixed command in SPSS 11.5 Syntax manual.

Table 2. Description of SPSS Mixed syntax

1.	GET FILE = 'C:/hsb12.sav' .	Opens a data file.
2.	MIXED mathach BY sector WITH meanses cses	The procedure (MIXED), the outcome variable (mathach), and factor (BY) and/or covariates (WITH) are specified on the first line. Note: the SPSS approach applies traditional ANOVA terminology in that categorical predictors are referenced as 'Factor' and listed after the BY term; where continuous predictors are referred to as 'Covariate' and are listed after the WITH term.
3.	/METHOD = REML	METHOD of estimation can be REML or ML. REML is the default option for SPSS and SAS. [See Singer & Willett (2003) p. 87 - 90 for a comparison of REML and ML.]
4.	/PRINT = SOLUTION TESTCOV	SOLUTION specifies that estimates of FIXED effects parameter estimates are printed, TESTCOV specifies that the covariance tests are included with random effects parameter estimates.
5.	/FIXED = meanses sector cses meanses*cses sector*cses | SSTYPE(3)	Predictor variables are specified after FIXED, SSTYPE(3) is the default Sums of Squares test of significance.
6.	/RANDOM = INTERCEPT cses | SUBJECT(school) COVTYPE(UN).	Variables entered after the RANDOM statement are allowed to vary at level 2. The nesting variable is listed in parentheses after SUBJECT, followed by specification of the Covariance matrix model, in this case the covariance matrix is modeled as unstructured (UN).

Examples of Multilevel Models Using SPSS Mixed Syntax

Data

High School and Beyond (HSB) data used in the multilevel examples can be obtained from the ICPSR web site at: http://www.icpsr.umich.edu/cgi/archive.prl?study=8443 or the student edition of HLM located at: http://www.ssicentral.com/other/hlmstu.htm. The HSB data file is included as part of the HLM example data set. Data analyzed in the growth model examples were obtained from: http://www.ats.ucla.edu/stat/examples/alda/ select Download under the SPSS to download the zip file aldaspss.zip, then find the data file opposites_pp.sav.

In an effort to provide a minimal of context the following discussion paraphrases work by Singer (1998). However the purpose of this paper is to demonstrate how to replicate the results of Singer (1998) using SPSS not to make a substantive contribution to the work of Singer (1998), nor any other research involving these particular data.

Example 1: The Unconditional Model.

```
MIXED  mathach
  /METHOD = REML
  /PRINT =  SOLUTION TESTCOV
  /FIXED = | SSTYPE(3)
  /RANDOM = INTERCEPT  | SUBJECT(school ) COVTYPE(UN).
```

This is an *unconditional model* because there are no predictors. As such it generally serves as a baseline by which to evaluate subsequent models.

First, the estimated overall School average on Mathach is 12.6 and is statistically significant. The 'Estimates of Covariance Parameters' section provides an estimate of the random effects of the model—i.e. variance in the intercepts and residual. In this case the estimated variance of the Intercepts = 8.06 and the estimated value of the residual is 39.14. The first thing to notice is that these variance components are statistically significant, although it is well known that these tests may not be reliable (Singer, 1998; Singer & Willett 2003). Also noteworthy is that the variance between-students is nearly 5 times that of the variance between-schools, indicating a substantial amount of variation is due to within school differences. This is not to imply that a multilevel model is not appropriate in this situation. The intraclass correlation can be estimated as 8.6 / 8.6 + 39.1 = .18, indicating that without appropriate modeling of the nonindependence component of these data inaccurate statistical tests will result (Singer, 1998) .

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	12.63697	.244393	156.64	51.707	.000	12.1542419	13.1197058

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		39.148321	.660644	59.258	.000	37.8746616	40.4648133
Intercept [subject = SCHOOL]	Variance	8.614024	1.078803	7.985	.000	6.7391217	11.0105479

Example 2: Conditional –Level 2 Predictor

```
MIXED  mathach  WITH meanses
  /METHOD = REML
  /PRINT =  SOLUTION TESTCOV
  /FIXED = meanses | SSTYPE(3)
  /RANDOM = INTERCEPT  | SUBJECT(school ) COVTYPE(UN).
```

The only difference in this equation and the first is the inclusion of **meanses** on the FIXED line. The estimate of the school mean math achievement is 12.65 when the value of the predictors (meanses) is zero. Because meanses is centered at the grand mean (it has a mean of 0), the predicted mathach for a school of average ses is 12.65. Accordingly, the estimated value of 5.86 for the fixed effect meanses indicates that for each 1 unit increase in meanses the expected value of mathach increases by 5.86 units. The statistical test for meanses indicates that the null hypothesis of no relationship between school SES and math achievement can be rejected.

Covariance parameters or RANDOM effects, when compared to the previous model, indicates that very little within school (or residual variance) has been explained (39.15 vs. 39.16). However a substantial proportion of the between school variance in means can be explained by meanses given the between school variation in means has been reduced from 8.61 to 2.69—a 69% reduction in unexplained variance between school mean mathach.

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	153.743	7180.2	.000
MEANSES	1	153.407	263.15	.000

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	12.6494	.14928	153.74	84.736	.000	12.3545303	12.9443404
MEANSES	5.86353	.36145	153.40	16.222	.000	5.1494606	6.5776163

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		39.157082	.660801	59.257	.000	37.8831195	40.4738864
Intercept [subject = SCHOOL]	Variance	2.6387080	.404338	6.526	.000	1.9541536	3.5630668

Model 3: Conditional Model – Centered Level 1 Predictor (Student)
In order to model the effect of ses relative to a particular school, the student level variable ses is centered by subtracting from in the meanses *within* the student's school.
The SPSS code (sans brackets) is: [COMPUTE cses = ses - meanses . EXECUTE .].

```
MIXED  mathach  WITH cses
  /METHOD = REML
  /PRINT =   SOLUTION TESTCOV
  /FIXED = cses | SSTYPE(3)
  /RANDOM = INTERCEPT  cses | SUBJECT(school ) COVTYPE(UN).
```

Notice the difference between this model and the previous one. In addition to including cses as a predictor on the FIXED line, it is also included in the RANDOM specification part of the model.

The Fixed Effects output indicates that the overall school mean for mathach is 12.65 after controlling for student ses. The estimated relationship between student ses and mathach is 2.19. Both of these parameters are statistically significant and one can conclude that there is a significant relationship between student ses and math achievement.

The covariance parameters estimates indicate how much the intercept and slope vary across schools. Variation in intercepts is 8.68, slopes 0.69, and the covariance between the two 0.05. Three points can be drawn from these estimates. 1) Schools vary significantly in average math achievement after controlling for student ses [cses]. 2) The relationship between cses and mathach varies significantly between schools at the .05 level. 3) There is little correlation in the *relationship* between cses and average school math achievement and school achievement. In other words the relationship is not stronger / weaker for schools with high average mathach.

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	156.751	2676.2	.000
CSES	1	155.218	292.40	.000

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	12.649338	.244513	156.75	51.733	.000	12.1663727	13.1323048
CSES	2.1931921	.128258	155.21	17.100	.000	1.9398341	2.4465501

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		36.700196	.6257440	58.650	.000	35.4940269	37.9473549
Intercept + CSES [subject = SCHOOL]	UN (1,1)	8.6816434	1.079625	8.041	.000	6.8037571	11.0778399
	UN (2,1)	.0507473	.4063926	.125	.901	-.7457676	.8472623
	UN (2,2)	.6939945	.2807858	2.472	.013	.3140257	1.5337226

Model 4: Conditional Model – Centered Level 1 and Level 2 Predictors

```
MIXED  mathach  BY sector WITH meanses cses
  /METHOD = REML
  /PRINT  = SOLUTION TESTCOV
  /FIXED  = meanses sector cses meanses*cses sector*cses  | SSTYPE(3)
  /RANDOM = INTERCEPT cses  | SUBJECT(school ) COVTYPE(UN).
```

This model includes a second level 2 predictor indicating school type. The categorical variable sector is coded as 0 for public and 1 for Catholic. Note that sector is listed after the BY statement, indicating that it is a factor or categorical variable.

All Fixed effects are significant. Because meanses and cses are both centered, the average mathach score for public schools is 12.22 and 13.33 for Catholic schools.
The interaction terms suggest that student ses is more important in schools with high average ses, and in public schools. The random effects however suggest that allowing the slopes cses to vary between schools is not useful. Subsequent analysis in which only intercepts are allowed to vary would be appropriate.

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	149.226	7884.19	.000
MEANSES	1	150.970	209.018	.000
SECTOR	1	149.600	15.769	.000
CSES	1	144.436	355.702	.000
MEANSES * CSES	1	160.562	12.080	.001
CSES(SECTOR)	1	143.353	46.923	.000

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	13.3302571	.2201540	141.627	60.550	.000	12.8950444	13.7654698
MEANSES	5.3391182	.3692988	150.970	14.457	.000	4.6094569	6.0687796
[SECTOR=0]	-1.2166722	.3063854	149.600	-3.971	.000	-1.8220739	-.6112706
[SECTOR=1]	0(a)	0	.	.	.	.	.
CSES	1.2961798	.1729351	147.671	7.495	.000	.9544326	1.6379269
MEANSES * CSES	1.0388706	.2989010	160.562	3.476	.001	.4485862	1.6291550
CSES([SECTOR=0])	1.6425829	.2397914	143.353	6.850	.000	1.1685990	2.1165667
CSES([SECTOR=1])	0(a)	0	.	.	.	.	.

a This parameter is set to zero because it is redundant.

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		36.7211286	.6261327	58.648	.000	35.5142106	37.9690627
Intercept + CSES [subject = SCHOOL]	UN (1,1)	2.3818588	.3717483	6.407	.000	1.7541424	3.2342021
	UN (2,1)	.1926034	.2045243	.942	.346	-.2082569	.5934637
	UN (2,2)	.1013798	.2138116	.474	.635	.0016246	6.3262882

Examples of Growth Models Using SPSS Mixed Syntax

Model 5. Unconditional Growth Model

```
MIXED  opp  WITH time
  /METHOD = REML
  /PRINT =  SOLUTION  TESTCOV
  /FIXED = time | SSTYP(3)
  /RANDOM = INTERCEPT time | SUBJECT(id ) COVTYPE(UN).
```

As usual, the initial focus is on the Fixed effects of the model. In this case the average intercept (starting point) is 164.37 and the average slope (change) is 26.96. Inspection of the random effects indicates that the residual variance is 159.48 and is statistically significant. The intercept and slope also have significant variance, indicating variance that is potentially explained by predictor variables.

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	34	721.653	.000
TIME	1	34	154.839	.000

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	164.3742857	6.1188486	34	26.864	.000	151.9392892	176.8092822
TIME	26.9600000	2.1666037	34	12.443	.000	22.5569316	31.3630684

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		159.4771429	26.9565572	5.916	.000	114.5035958	222.1149381
Intercept + TIME [subject = ID]	UN (1,1)	1198.7767899	318.3809670	3.765	.000	712.3103292	2017.4715052
	UN (2,1)	-179.2555630	88.9634162	-2.015	.044	-353.6206548	-4.8904712
	UN (2,2)	132.4005714	40.2106963	3.293	.001	73.0089022	240.1064908

Model 6. Conditional Growth Model

```
MIXED  opp  WITH time ccog
  /METHOD = REML
  /PRINT =  SOLUTION  TESTCOV
  /FIXED = time ccog time*ccog | SSTYP(3)
  /RANDOM = INTERCEPT time | SUBJECT(id ) COVTYPE(UN).
```

This model includes the covariate ccog. The interpretation of the intercept and slope is similar to the unconditional model, except now the term ‘controlling for ccog’ precedes their discussion. The effect of ccog is 0.43, thus individuals who differ by 1 on ccog differ by .43 in their rate of change. Perhaps the most interesting aspect of this model is that the variance component for growth rate went from 132.40 to 107.25, an approximately 19% reduction in variation.

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	33	701.505	.000
TIME	1	33	182.827	.000
CCOG	1	33	.051	.823
TIME * CCOG	1	33	7.146	.012

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	164.3742906	6.2060954	33	26.486	.000	151.7478945	177.0006866
TIME	26.9599814	1.9938808	33	13.521	.000	22.9034005	31.0165624
CCOG	-.1135527	.5040119	33	-.225	.823	-1.1389726	.9118672
TIME * CCOG	.4328577	.1619278	33	2.673	.012	.1034131	.7623024

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Residual		159.4771429	26.9565572	5.916	.000	114.5035958	222.1149381
Intercept + TIME [subject = ID]	UN (1,1)	1236.4127057	332.4021783	3.720	.000	730.0002411	2094.1313343
	UN (2,1)	-178.2332472	85.4297778	-2.086	.037	-345.6725348	-10.7939596
	UN (2,2)	107.2491911	34.6767044	3.093	.002	56.9084074	202.1210841

Model 7. Conditional Growth Model With Repeated Statement

```
MIXED  opp  BY wave WITH time ccog
  /METHOD = REML
  /PRINT =  SOLUTION  TESTCOV
  /FIXED = time ccog time*ccog | SSTYP(3)
  /RANDOM = INTERCEPT time | SUBJECT(id) COVTYPE(UN)
  /REPEATED = wave  | SUBJECT(id ) COVTYPE(AR1).
```

Inclusion of the REPEATED subcommand differentiates this model from the previous. The REPEATED statement models the *within* subject variance, as opposed to the RANDOM statement which models the *between* subject variance. The REPEATED statement is necessary models if there is a meaningful relationship across measurements within persons, as indicated by the specification of the COVTYPE(AR1). AR1 represents autoregressive relationship with a lag of 1. In other words, each person's response is strongly correlated to their previous response. The wave variable is numerically equal to time, however by including wave on the BY statement SPSS treats is like a categorical variable (whereas time is considered as continuous when specified after the WITH statement. The variable used on the REPEATED statement is not required to be categorical.

Type III Tests of Fixed Effects

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	32.990	703.525	.000
TIME	1	32.531	185.162	.000
CCOG	1	32.990	.060	.808
TIME * CCOG	1	32.531	7.362	.011

Estimates of Fixed Effects

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	164.4227269	6.1990069	32.990	26.524	.000	151.8106078	177.0348459
TIME	26.9081610	1.9774614	32.531	13.607	.000	22.8827827	30.9335394
CCOG	-.1233819	.5034362	32.990	-.245	.808	-1.1476423	.9008786
TIME * CCOG	.4357299	.1605944	32.531	2.713	.011	.1088193	.7626406

Estimates of Covariance Parameters

Parameter		Estimate	Std. Error	Wald Z	Sig.
Repeated Measures	AR1 diagonal	141.3693337	36.3484257	3.889	.000
	AR1 rho	-.1369391	.2588982	-.529	.597
Intercept + TIME [subject = ID]	UN (1,1)	1258.1004051	333.2509094	3.775	.000
	UN (2,1)	-182.4125474	84.5530168	-2.157	.031
	UN (2,2)	110.9420681	34.5303148	3.213	.001

Evaluation of complex models such as this one is usually done by comparing it to simpler models. However, it is clear that the AR1 rho element of the covariance matrix, which represents the inclusion of the AR1 structure on the REPEATED statement, is not distinguishable from 0 and is of no utility in the model.

References

Singer, J. (1998). Using SAS Proc Mixed to fit multilevel Models, Hierarchical Models, and Individual Growth Curves. Journal of Educational and Behavioral Statistics. Vol 24, No 4, pp 323-355. (also available at: http://www.gse.harvard.edu/~faculty/singer/)

Snijders, T. & Bosker, R. (2000). Multilevel Analysis: An introduction to basic and advanced multilevel modeling. London:Sage.

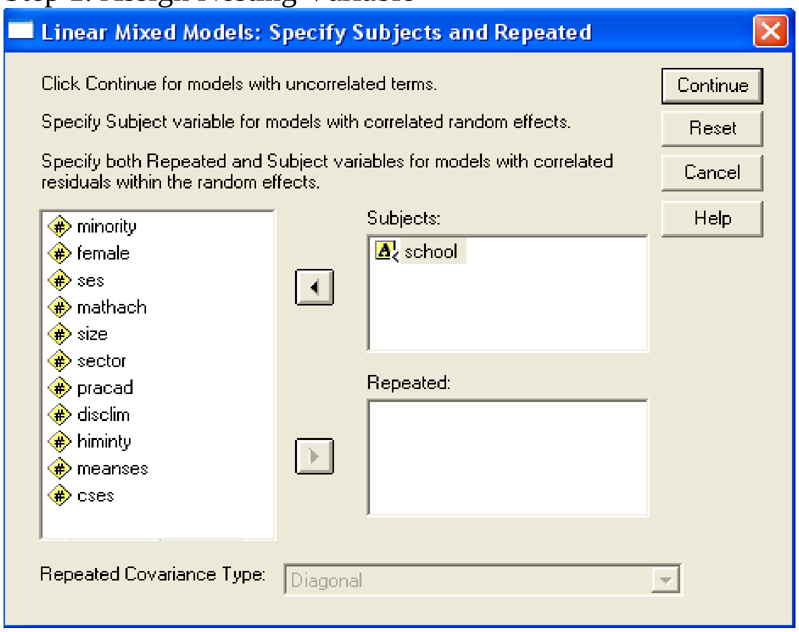
Singer, J. & Willett, J. (2003). Applied Longitudinal Data Analysis. Oxford: Oxford University Press.

Appendix

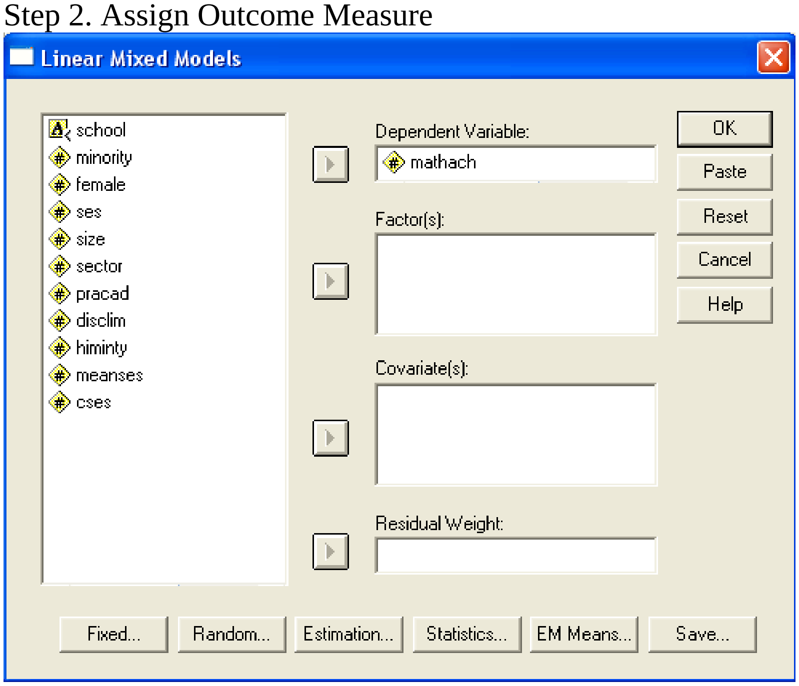
Begin the process of building SPSS syntax via menu options by selecting Analyze > Mixed Models > Linear. At the next screen specify the Subjects (nesting variable) and Repeated variables (in the case of growth modeling). In this case the nesting factor is the variable School., so School is moved to the Subjects window.

Example 1. Unconditional Model (Model 1).

Step 1. Assign Nesting Variable

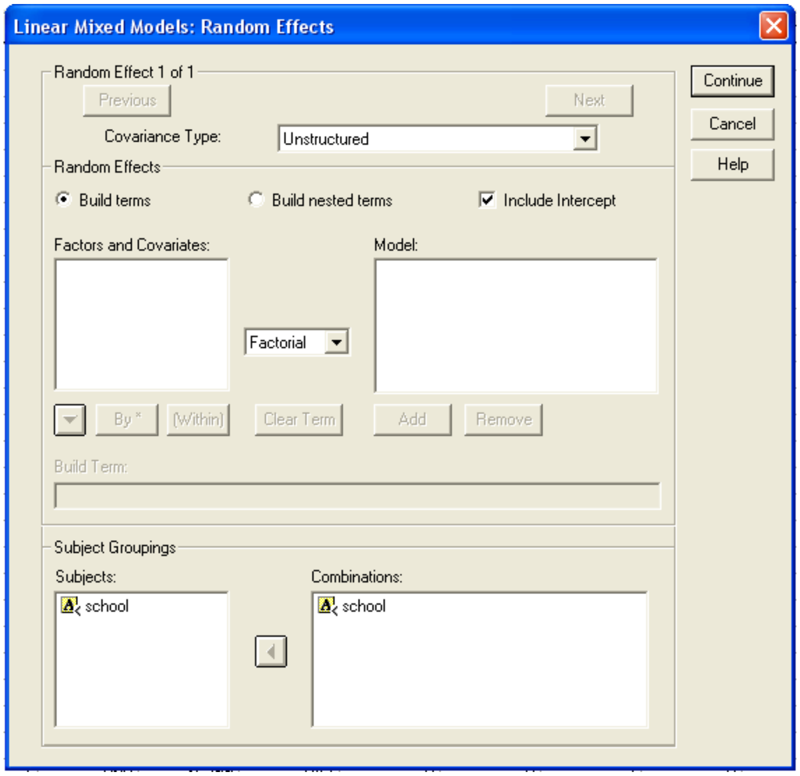


Step 2. Assign Outcome Measure

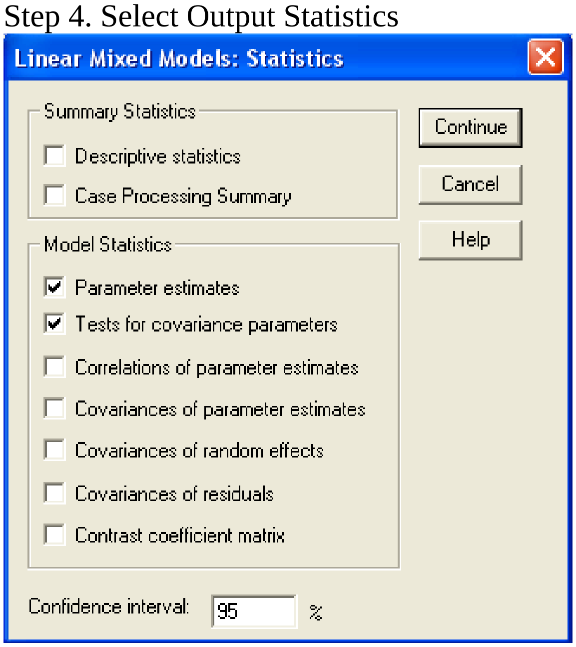


After the variables are specified it is necessary to define the model by selecting the appropriate Statistics (output options), Estimation (method, iterations, etc), Fixed (fixed effects), Random (random effects), and Save options.

Step 3. Select Random effects variables, Covariance Type, Intercept, Subject Grouping Combinations.



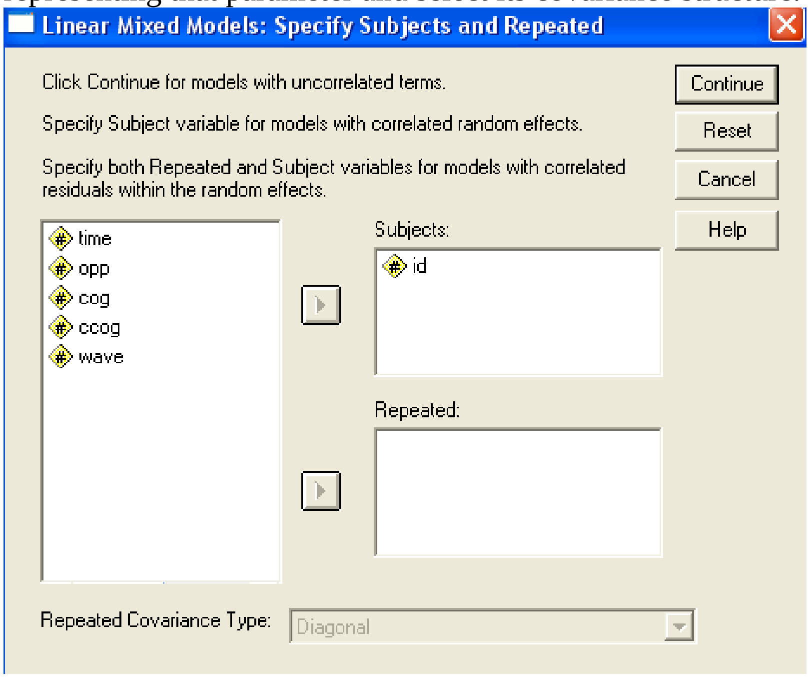
Step 4. Select Output Statistics



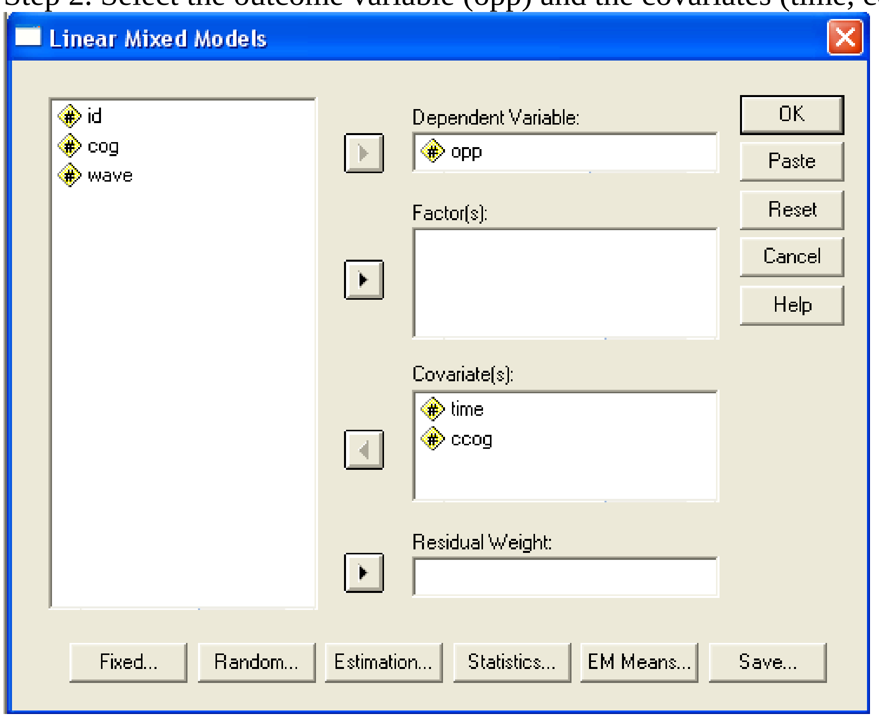
This is all that is required to create the basic unconditional model. Select either Ok or Paste to execute this analysis.

Example 2: Conditional Growth Modeling via Menu Options (Model 6).

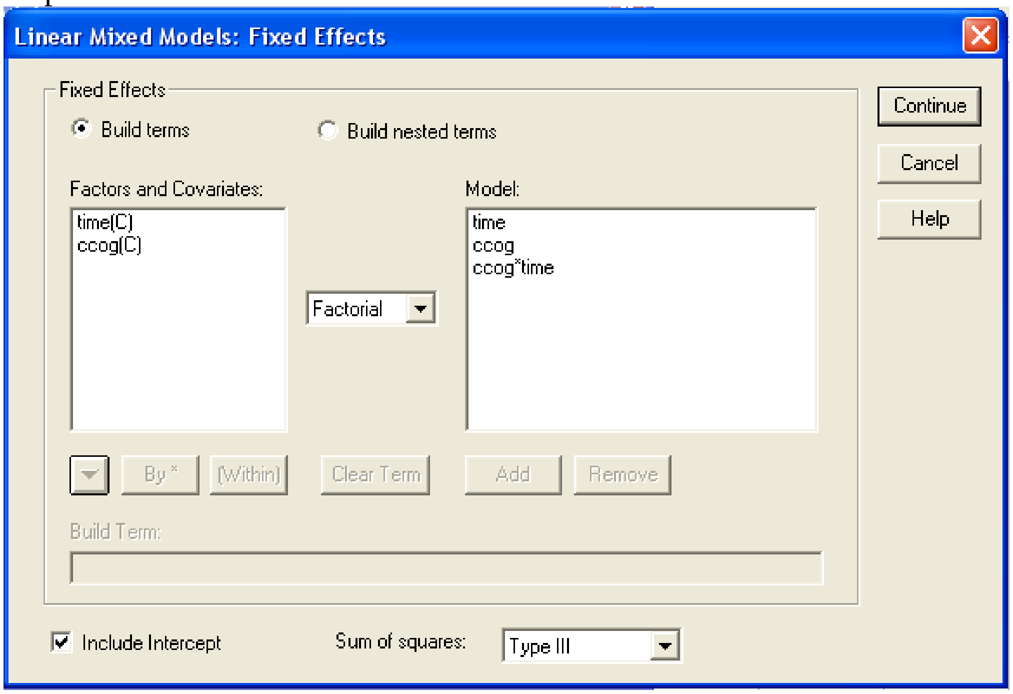
Step 1. Select the nesting variable. If REPEATED option is used then select the variable representing that parameter and select its covariance structure.



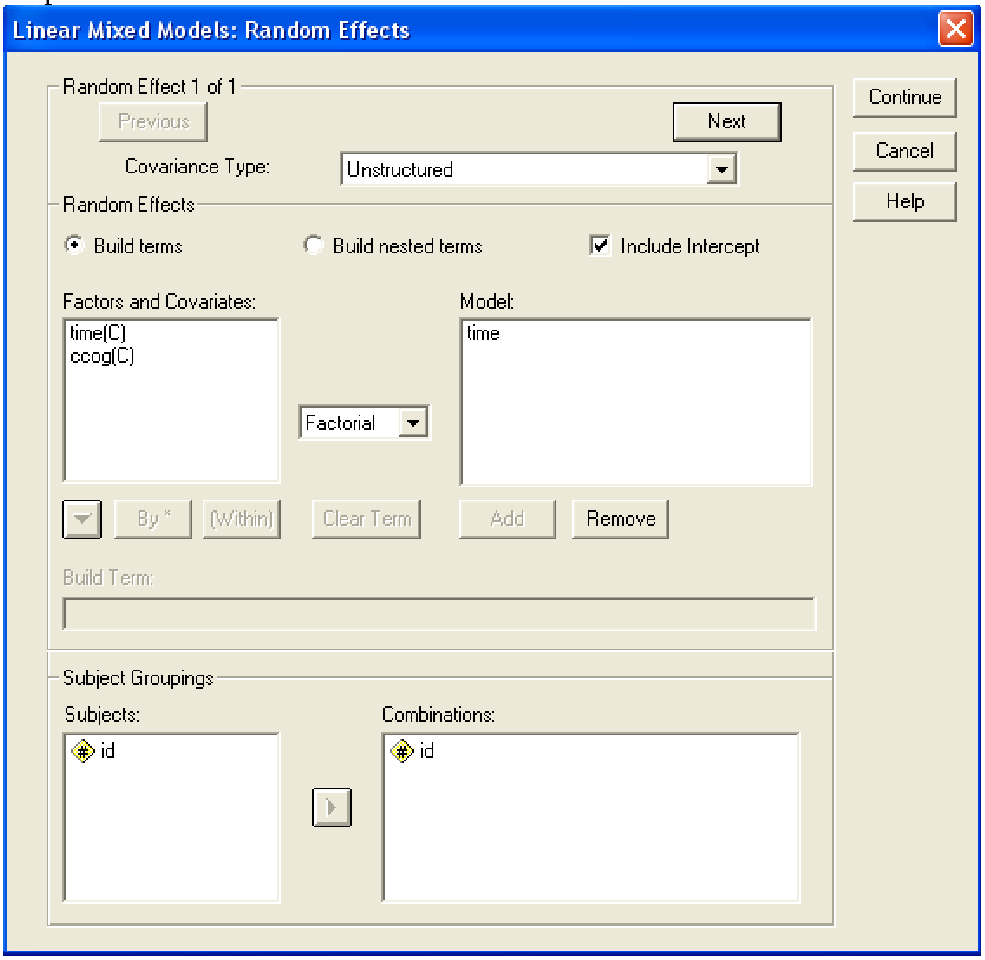
Step 2. Select the outcome variable (opp) and the covariates (time, ccog)



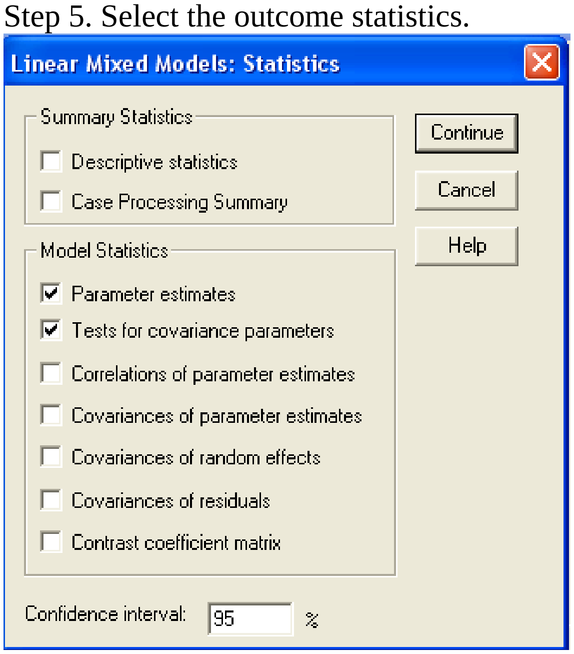
Step 3. Build the FIXED effects term.



Step 4. Build the RANDOM effects term.



Step 5. Select the outcome statistics.



Select either Ok or Paste to execute this analysis.