

Statistical Computing Workshop: Using the SPSS Mixed Command

Introduction

The purpose of this workshop is to show the use of the **mixed** command in SPSS. Although it has many uses, the **mixed** command is most commonly used for running linear mixed effects models (i.e., models that have both fixed and random effects). Such models are often called multilevel models. Because the purpose of this workshop is to show the use of the **mixed** command, rather than to teach about multilevel models in general, many topics important to multilevel modeling will be mentioned but not discussed in detail. References will be provided so that those interested in these topics can find additional information. The **mixed** command in SPSS is used to run linear regression models, including mixed effects models.

When most people think of linear regression, they think of ordinary least squares (OLS) regression. In this type of regression, the outcome variable is continuous, and the predictor variables can be continuous, categorical, or both. How would you run a linear regression model in SPSS? Perhaps you would use either the **regression** command or the **glm** command.

```
regression
/dep = write
/method = enter read female.
```

```
glm write with read female
/print = parameter.
```

Here is the same model run with the **mixed** command.

```
mixed write with read female
/fixed read female
/print = solution.
```

The SPSS keyword **with** is used with both the **glm** and the **mixed** commands to indicate that the two predictor variables, **read** and **female**, are to be treated as continuous. The **print** subcommand is used to have the parameter estimates included in the output (although the options used on the subcommand are different).

Obviously, there are already several ways to run an OLS regression in SPSS, so what else can the **mixed** command do? It can run linear models with clustered data. Remember that one of the most important assumptions of OLS regression is that the observations are independent. From a practical stand point, this means that each observation is represented on one and only one row in the data file. The effects of violating this assumption depend on how the assumption is violated. If this assumption is violated by having clustering in the data, the standard errors around the point estimates will be underestimated, and false alarms will be more likely. Please see Harrell (2015) and Scariano and Davenport (1987) for explanations and examples.

Repeated measures

How do data become clustered? One way is to collect data on the same subjects over time. For example, a researcher might measure subjects' blood pressure at three different time points. Let's take a moment to look at our example data set. We have the outcome at three time points and two categorical predictors, diet (**diet**) and exercise type (**exertype**). Because multiple observations are made on the same subjects, each data point is related to all of the other data points collected from that subject. Technically, we say that the errors within subjects are correlated. Because the observations are not independent, we cannot use the **regression** command, or the standard errors will be biased, and hence, the test statistics and p-values will be inaccurate. So what procedure can be used? A common choice is the **glm** command with the **wsfactor** subcommand. We are going to start with an example of repeated measures ANOVA because, like OLS regression, most researchers are familiar with this type of analysis. In order to use the **glm** command for a repeated measures ANOVA, the data must be structured in wide format.

```
data list free / id exertype diet time1 time2 time3.
begin data.
1 1 1 85 85 88
2 1 1 90 92 93
```

```

3    1    1    97    97    94
4    1    1    80    82    83
5    1    1    91    92    91
6    1    2    83    83    84
7    1    2    87    88    90
8    1    2    92    94    95
9    1    2    97    99    96
10   1    2    100   97    100
11   2    1    86    86    84
12   2    1    93    103   104
13   2    1    90    92    93
14   2    1    95    96    100
15   2    1    89    96    95
16   2    2    84    86    89
17   2    2    103   109    90
18   2    2    92    96    101
19   2    2    97    98    100
20   2    2    102   104    103
21   3    1    93    98    110
22   3    1    98    104   112
23   3    1    98    105    99
24   3    1    87    132   120
25   3    1    94    110   116
26   3    2    95    126   143
27   3    2    100   126   140
28   3    2    103   124   140
29   3    2    94    135   130
30   3    2    99    111   150
end data.

```

Let's say that the hypothesis to be tested is that **diet** affects blood pressure. Blood pressure readings from three times are contained in the variables **time1**, **time2** and **time3**. The **glm** syntax for this would be:

```

glm time1 time2 time3 by diet
/wsfactor = time 3
/wsdesign = time
/design = diet.

```

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

	Approx. Chi-Square	df	Asymp. Sig.	Greenhouse-Geisser	Epsilon ^b	Lower bound
1	1.111	2	.578	.500	.500	.500

Within Subjects Effect	Mauchly's W	Square	df	Sig.	Geisser	Huynh-Feldt	Lower-bound
time	.673	10.679	2	.005	.754	.816	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

- a. Design: Intercept + diet
Within Subjects Design: time
- b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
time	Sphericity Assumed	2066.600	2	1033.300	11.808	.000
	Greenhouse-Geisser	2066.600	1.508	1370.842	11.808	.000
	Huynh-Feldt	2066.600	1.632	1266.577	11.808	.000
	Lower-bound	2066.600	1.000	2066.600	11.808	.002
time * diet	Sphericity Assumed	192.822	2	96.411	1.102	.339
	Greenhouse-Geisser	192.822	1.508	127.905	1.102	.326
	Huynh-Feldt	192.822	1.632	118.177	1.102	.330
	Lower-bound	192.822	1.000	192.822	1.102	.303
Error(time)	Sphericity Assumed	4900.578	56	87.510		
	Greenhouse-Geisser	4900.578	42.211	116.097		
	Huynh-Feldt	4900.578	45.686	107.267		
	Lower-bound	4900.578	28.000	175.021		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	time	Type III Sum of Squares	df	Mean Square	F	Sig.
time	Linear	1915.350	1	1915.350	14.202	.001
	Quadratic	151.250	1	151.250	3.767	.062
time * diet	Linear	190.817	1	190.817	1.415	.244
	Quadratic	2.006	1	2.006	.050	.825
Error(time)	Linear	3776.333	28	134.869		
	Quadratic	1124.244	28	40.152		

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results2.png)

[content/uploads/2018/02/results2.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results2.png)

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	894608.100	1	894608.100	2231.137	.000
diet	1261.878	1	1261.878	3.147	.087
Error	11227.022	28	400.965		

[https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results3.png\).](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results3.png)

General Linear Model

Within-Subjects
Factors

Measure: MEASURE_1	
	Dependent Variable
1	time1
2	time2
3	time3

Between-Subjects Factors

	N	
diet	1.00	15
	2.00	15

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.
time	Pillai's Trace	.357	7.479 ^b	2.000	27.000	.003
	Wilks' Lambda	.643	7.479 ^b	2.000	27.000	.003
	Hotelling's Trace	.554	7.479 ^b	2.000	27.000	.003
	Roy's Largest Root	.554	7.479 ^b	2.000	27.000	.003
time * diet	Pillai's Trace	.056	.801 ^b	2.000	27.000	.459
	Wilks' Lambda	.944	.801 ^b	2.000	27.000	.459
	Hotelling's Trace	.059	.801 ^b	2.000	27.000	.459
	Roy's Largest Root	.059	.801 ^b	2.000	27.000	.459

a. Design: Intercept + diet
Within Subjects Design: time

b. Exact statistic

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results1.png>

We can see that **time** is statistically significant, but **diet** and the **time** by **diet** interaction is not. Wait a minute – **time** by **diet** interaction? Where did we specify that in our **glm** syntax? An interaction of the variable listed on the **wsfactor** subcommand and the predictor variable(s) is included in the model by default, at least for the multivariate tests. From the SPSS documentation for the GLM: Repeated Measures entry we learn:

“A repeated measures analysis includes a within-subjects design describing the model to be tested with the within-subjects factors, as well as the usual between-subjects design describing the effects to be tested with between-subjects factors. The default for the within-subjects factors design is a full factorial model which includes the main within-subjects factor effects and all their interaction effects.”

For some analyses, this may be appropriate, but what if you don't want this interaction in model? The answer is that you must use a different command, because there is no way to override this default.

Relationship between time points

Now let's think about the relationship between these repeated observations. We know that they are not independent, but how, exactly, are they related? We could imagine that the relationship was the same across time, so that it was either set to a specific constant (such as 0) or estimated as a single value. At the other extreme, we could assume that every value needed to be estimated. Another possibility is that the relationship between the time points decreases as the distance between the time points increases; in other words, time points that are close together are more closely related than time points that are further apart.

	time1	time2	time3
time1			
time2			
time3			

So what did SPSS use? For the univariate results, SPSS used compound symmetry (CS). This structure estimates one value for the variances and another value for the covariances. For the multivariate results an unstructured (UN) covariance structure was used, meaning that every value in the variance-covariance matrix was estimated. These defaults are used by all statistical software packages for repeated measures analyses. This is default that cannot be altered in the **glm** command, even if it does not do a good job of representing the data.

Missing data

Let's consider one more important point about these data.

id	exertype	diet	time1	time2	time3
1	1	1	85	85	88
2	1	1	90	92	93
3	1	1	97	97	94
4	1	1	80	82	83
5	1	1	91	92	91
...					

What if we had been unable to measure S3's blood pressure at time 3? Now suppose that we were also unable to measure S2's blood pressure at time 1? SPSS (and all other general-use statistical software) uses listwise deletion by default. This means that even a little bit of missing data could be a big problem. Another limitation of the repeated measures ANOVA analysis is that all of the predictors must be at the level of the person; we cannot include any predictors that vary by time (Harrell, 2015).

Reshaping the data from wide to long

Since this is a workshop on the SPSS **mixed** command, let's find out how the **mixed** command can help us out with these issues. We will start by running our model with the **mixed** command. Before we can do that, however, we will first need to restructure the data. The **glm** command requires the data to be in wide form, but **mixed** command requires the data to be in long form. We can use the **varstocases** command to restructure the data from wide to long.

```
varstocases
/make bp from time1 time2 time3
/index = time.
```

Variables to Cases

Generated Variables	
Name	Label
time	<none>
bp	<none>

Processing Statistics	
Variables In	6
Variables Out	5

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results4.png>

To be clear on what just happened, let's look at the data before and after the restructuring:

Data in wide form

```
id exertype diet time1 time2 time3
1 1 1 85 85 88
2 1 1 90 92 93
3 1 1 97 97 94
4 1 1 80 82 83
5 1 1 91 92 91
```

Data in long form

```
id exertype diet time bp
1 1 1 1 85
1 1 1 2 85
1 1 1 3 88
2 1 1 1 90
2 1 1 2 92
2 1 1 3 93
3 1 1 1 97
3 1 1 2 97
3 1 1 3 94
4 1 1 1 80
4 1 1 2 82
4 1 1 3 83
5 1 1 1 91
5 1 1 2 92
5 1 1 3 91
```

Data in wide format are also known as multivariate, because there are multiple dependent variables (DVs). Data in long format are also known as univariate, because there is a single DV. So we have gone from having the DV spread across multiple columns to having the DV in a single column.

An example using the mixed command with the repeated subcommand

Now let's get back to the **mixed** command, using, of course, the data in long format. Now it is easy to exclude unwanted interaction terms by simply not including them on the **fixed** subcommand. We can change the covariance structure by altering the structure given in the **covtype** option on the **repeated** subcommand.

```
mixed bp by time diet
/fixd diet time
/repeated time | subject(id) covtype(ar1)
/print = solution.
```

Mixed Model Analysis

Model Dimension ^a				
Number of	Covariance	Number of	Subiaect	Number of

		Number of Levels	Estimate Structure	Number of Parameters	Number of Variables	Number of Subjects
Fixed Effects	Intercept	1		1		
	diet	2		1		
	time	3		2		
Repeated Effects	time	3	First-Order Autoregressive	2	id	30
Total		9		6		

a. Dependent Variable: bp.

Information Criteria^a

-2 Restricted Log Likelihood	673.531
Akaike's Information Criterion (AIC)	677.531
Hurvich and Tsai's Criterion (AICC)	677.676
Bozdogan's Criterion (CAIC)	684.440
Schwarz's Bayesian Criterion (BIC)	682.440

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: bp.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results5.png)

[content/uploads/2018/02/results5.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results5.png)

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	29.854	2146.652	.000
diet	1	29.676	3.188	.084
time	2	57.584	10.661	.000

a. Dependent Variable: bp.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	108.238379	3.298843	40.135	32.811	.000	101.571866	114.904892
[diet=1.00]	-7.610091	4.261899	29.676	-1.786	.084	-16.318036	1.097853
[diet=2.00]	0 ^b	0
[time=1]	-11.300000	2.646649	74.125	-4.270	.000	-16.573414	-6.026586
[time=2]	-2.900000	2.048589	57.237	-1.416	.162	-7.001860	1.201860
[time=3]	0 ^b	0

a. Dependent Variable: bp.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Repeated Measures		
AR1 diagonal	190.242532	38.038483
AR1 rho	.669103	.078563

a. Dependent Variable: bp.

[https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results6.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results6.png)

How are missing data handled? Listwise deletion is still used, but the effect on the analysis is different.

```
id exertype diet time bp
```

```
1 1 1 1 85
```

```
1 1 1 2 85
```

```

1 1 1 3 88
2 1 1 1 90
2 1 1 2 92
2 1 1 3 93
3 1 1 1 97
3 1 1 2 97
3 1 1 3 94
4 1 1 1 80
4 1 1 2 82
4 1 1 3 83
5 1 1 1 91
5 1 1 2 92
5 1 1 3 91

```

Even if datum from one time point is missing, the data from the other time points are still included in the analysis. This illustrates an important point: Subjects do not need to have the same number of observations. In other words, subjects can have different numbers of rows of data. As we will see in some examples a little later on in this workshop, adding time-varying predictors variables is as simple as listing them on the **mixed** command and on the **fixed** subcommand.

Let's consider two more important points before leaving this analysis. First, with a repeated measures ANOVA, the time points are to be equally spaced. Secondly, everyone is to be measured at those exact times. In our example, subjects' blood pressure was measured at 15 minute intervals. In carefully controlled experiments, this isn't usually a problem. In other research situation, however, this isn't feasible. Mixed effects models do not require that subjects be measured at the same intervals.

The take-away message is not that repeated measures ANOVAs are bad or flawed, but rather that repeated measures ANOVAs are a limited set of multilevel models. Indeed, mixed effect analyses are themselves a limited case of another type of analysis. On page 145 of Harrell (2015), there is a very nice table that lists the different ways that repeated measures data can be analyzed and notes the features and limitations of each of these options.

Clustered data

Let's return to the question of how data might be clustered and consider another example. Instead of observations clustered within a person, we could have people clustered in families, patients clustered in hospitals or students clustered in classrooms. In our next example dataset, we will have students clustered, or nested, in classrooms. We will use the [hsbdemo](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsbdemo.sav) (<https://stats.idre.ucla.edu/wp-content/uploads/2016/02/hsbdemo.sav>) dataset in our examples. The variable **cid** is the classroom identifier, and the variable **id** is the student identifier.

Let's look at the first 15 observations for some of the variables that will be used in the upcoming examples.

```
sort cases by cid id.
```

```
list cid id write read female prog
/cases from 1 to 15.
```

```

cid  id write read female prog
  1   1   44   34     1     3
  1   2   41   39     1     3
  1  15   39   39     0     3
  1  45   35   34     1     3
  1  51   36   42     1     1
  1  53   37   34     0     3
  1  67   37   37     0     3
  1 108   33   34     0     1
  1 133   31   50     0     3
  1 153   31   39     0     3
  1 164   36   31     0     3
  2  11   46   34     0     2
  2  16   31   47     0     3
  2  19   46   28     1     1
  2  89   35   35     1     3

```

```
Number of cases read: 15    Number of cases listed: 15
```

An example of mixed without accounting for clustering

Suppose we wanted to build a linear regression model regressing **write** on **read** and **female**. Here is the **mixed** syntax to do that. (This is the same OLS regression model from earlier in the workshop.)

```

mixed write with read female
/fixed read female
/print = solution.

```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	197	55.562	.000
read	1	197	131.302	.000
female	1	197	29.265	.000

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	20.228368	2.713756	197	7.454	.000	14.876626	25.580111
read	.565887	.049385	197	11.459	.000	.468496	.663278
female	5.486894	1.014261	197	5.410	.000	3.486690	7.487098

a. Dependent Variable: writing score.

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	50.875908	5.126182

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results7.png>

Looking at the section of the output called Fixed Effects, we see two tables. The first table, the Type III Tests of Fixed Effects, gives the overall statistical significance for each of the predictors in the model. Because both of the predictors in our model have only one degree of freedom, this information is redundant with the information in the next table. However, if we had included categorical predictor variables in our model that had more than two levels, this table would be very useful, because it would give the multi degree of freedom test of the those variables. We will see an example of this later on in this workshop. The table called Estimates of Fixed Effects gives the coefficients, standard errors, degrees of freedom, test statistic, significance, and lower and upper confidence intervals for the predictors. It is usually the table that everyone hurries to once they have run a model. In this output, we can see that both **read** and **female** are statistically significant predictors. The coefficients, listed in the column called Estimates, are interpreted just like coefficients from an OLS regression. For example, for a one-unit increase in **read**, the expected value of **write** increases by 0.565887 unit. In the section of the output called Covariance Parameters, we see the estimate of the residual, which is 50.875908.

An example of the mixed command accounting for clustering

Now let's run this same model but account for the clustering of students in classrooms. One new subcommand has been added to account for the clustering of students in classrooms, the **random** subcommand. The SPSS keyword **intercept** has been included on this subcommand to specify a random intercept model. This type of model is commonly used to account for clustering in data.

Let's mention a few important points before we actually run the model. First, the lowest level of a multilevel model is called level 1. The outcome variable must be at level 1. A multilevel model must have at least two levels, and in our example here, the model only has two levels, so level 2 is the highest level. Predictor variables can be at any level of the model, and SPSS (and almost all statistical software packages) will automatically detect the level of each predictor; there is no need to provide that information. How does the software detect the level of each variable? If the variable changes within the level 2 identifier, then it is a level 1 variable. If the variable is constant within the level 2 identifier, then it is a level 2 variable. This means that all multilevel datasets must have a level 2 identifier (and an identifier for all levels above level 2, if there are any).

```
mixed write with read female
/fixed read female
/random intercept | subject(cid)
/print = solution.
```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	112.183	222.755	.000
read	1	195.291	8.515	.004
female	1	176.893	30.827	.000

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	60.612280	4.061133	112.183	14.925	.000	52.565809	68.658751
read	-.188057	.064447	195.291	-2.918	.004	-.315159	-.060956
female	3.897775	.702026	176.893	5.552	.000	2.512351	5.283200

a. Dependent Variable: writing score.

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	21.651709	2.311661
Intercept [subject = cid] Variance	91.667382	32.593592

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results8.png>

Let's look at the section of the output called Fixed Effects. As before, we can see in the table called Type III Tests of Fixed Effects that both **read** and **female** are statistically significant, although the F values are different. Looking at the table called Estimates of Fixed Effects, we can see that the coefficients are very different. The coefficient for **read** is now negative, such that for a one-unit increase in read, the expected value of **write** decreases by 0.188057.

Looking at the section called Covariance Parameters, we can see that another line is included in the table called Estimates of Covariance Parameter. The estimate of the level 1 residual is given on the first line as 21.651709. On the next line, we have the estimate of the variance around the grand intercept. The value is 91.667382. This value, like the value of the level 1 residuals, is rarely of interest and is not interpreted, although it is usually reported.

The binary variable **female** (0 = male, 1 = female) is treated as a continuous predictor variable in the analysis above because it is included after the SPSS keyword **with**, but it could be treated as a categorical variable as well. To do this, we can include it after the SPSS keyword **by**. When a predictor variable is specified as categorical, SPSS will by default use the highest numbered category as the reference group. This means that if a variable is coded as 0 and 1, the group coded 1 will be the reference group. Hence, the sign of the coefficient for the binary predictor will be different when it follows the SPSS keyword **with** than when it follows the SPSS keyword **by**. The intercept will also be different. There is no way to change the reference group of a categorical predictor variable in the **mixed** command; the only way to change the reference group is to create a new variable with the categories ordered differently. If this is needed, be certain to label the new variable and its values very carefully, so that there is no confusion with the original variable.

```
mixed write with read by female
/fixed read female
/random intercept | subject(cid)
/print = solution.
```

Fixed Effects

Type III Tests of Fixed Effects^a

	Denominator	
--	-------------	--

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	109.239	243.415	.000
read	1	195.291	8.515	.004
female	1	176.893	30.827	.000

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	64.510055	3.988960	108.300	16.172	.000	56.603494	72.416617
read	-.188057	.064447	195.291	-2.918	.004	-.315159	-.060956
[female=0]	-3.897775	.702026	176.893	-5.552	.000	-5.283200	-2.512351
[female=1]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	21.651709	2.311661
Intercept [subject = cid]	Variance	91.667382
		32.593592

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results9.png>

Notice that the values in the table called Type III Tests of Fixed Effects are the same as in the previous analysis. However, the coefficients for the intercept and **female** are different, although the coefficient for **read** is the same. The last line of the table is for **female** = 1, and the coefficient is 0. The footnote reminds us why: "The parameter is set to 0 because it is redundant." In other words, for a categorical predictor variable with only two levels, there can be only one coefficient. The fact that the estimated coefficient is for **female** = 0 reminds us that the reference group for **female** is 1, rather than 0. The values in the table called Estimates of Covariance Parameters are the same as the previous analysis.

When a predictor variable is specified as categorical, it can also be specified on the **emmeans** subcommand. The **emmeans** subcommand is used to get estimated marginal means, which can be thought of as a type of descriptive statistic that is based on the model. Estimated marginal means can help researchers better understand their results.

```
mixed write with read by female
/fixed read female
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female).
```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Sum of Squares	df	Mean Square	F	Sig.
--------	----------------	----	-------------	---	------

Source	Numerator df	df	F	Sig.
Intercept	1	109.239	243.415	.000
read	1	195.291	8.515	.004
female	1	176.893	30.827	.000

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	64.510055	3.988960	108.300	16.172	.000	56.603494	72.416617
read	-.188057	.064447	195.291	-2.918	.004	-.315159	-.060956
[female=0]	-3.897775	.702026	176.893	-5.552	.000	-5.283200	-2.512351
[female=1]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	21.651709	2.311661
Intercept [subject = cid]	Variance	91.667382

a. Dependent Variable: writing score.

Estimated Marginal Means

female^a

female	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
male	50.790 ^b	2.201	17.681	46.161	55.419
female	54.688 ^b	2.189	17.313	50.076	59.300

a. Dependent Variable: writing score.

b. Covariates appearing in the model are evaluated at the following values: reading score = 52.23.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results10.png>

Notice that the parameter for **female** = 0 is -3.89775. Now look at the values in the table called Estimated Marginal Means at the end of the output. The mean for males is 50.790 and the mean for females is 54.688. What value do you get if you subtract the mean for the females from the mean for the males? We have seen that binary predictor variables can be modeled as either continuous or categorical, but how should one choose which way to model such variables? In some sense, it does not really matter, because the models are equivalent. Notice that for both models, the -2 restricted log likelihood is 1253.994 and, in fact, all of the information in the table Information Criteria is exactly the same. This is also true of the table Type III Tests of Fixed Effects and the table Estimates of Covariance Parameters. To be clear, there is no "right" or "wrong" way to model a binary predictor: either way is OK. The decision on how to model a binary predictor (as continuous or categorical) in SPSS should depend on which group you want to have as the reference group and whether or not the variable is needed on the **emmeans** subcommand.

Adding more predictors

Now let's add another continuous predictor, **socst**, and another categorical predictor, **prog**, to the model. Notice that the SPSS keywords **with** and **by** are only specified once each; all of the continuous predictors are listed after the SPSS keyword **with** and all of the categorical predictor variables are listed after the SPSS keyword **by**. It does not matter if **with** or **by** is specified first, only that they are specified only once, if specified at all. Notice also that more than one **emmeans** subcommand can be specified.

```
mixed write with read socst by female prog
/fixed read socst female prog
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female)
/emmeans = tables(prog).
```

Fixed Effects

Type III Tests of Fixed Effects^a

Denominator	
-------------	--

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	128.488	185.350	.000
read	1	191.275	12.270	.001
socst	1	174.935	10.472	.001
female	1	173.174	30.498	.000
prog	2	173.250	.137	.872

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	59.006813	4.162933	126.041	14.174	.000	50.768516	67.245110
read	-.226546	.064674	191.275	-3.503	.001	-.354111	-.098981
socst	.140445	.043401	174.935	3.236	.001	.054788	.226102
[female=0]	-3.818242	.691398	173.174	-5.522	.000	-5.182894	-2.453589
[female=1]	0 ^b	0
[prog=1]	-.203334	1.011668	172.956	-.201	.841	-2.200139	1.793470
[prog=2]	.265629	.958305	173.634	.277	.782	-1.625798	2.157056
[prog=3]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	20.972557	2.263764
Intercept [subject= cid] Variance	78.348282	28.517896

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results11.png>

Estimated Marginal Means

1. female^a

female	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
male	50.737 ^b	2.047	17.306	46.424	55.050
female	54.555 ^b	2.036	16.920	50.259	58.851

a. Dependent Variable: writing score.

b. Covariates appearing in the model are evaluated at the following values: reading score = 52.23, social studies score = 52.41.

2. type of program^a

type of program	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
general	52.422 ^b	2.106	19.402	48.020	56.824
academic	52.891 ^b	2.040	17.073	48.588	57.194
vocation	52.625 ^b	2.113	19.662	48.212	57.038

a. Dependent Variable: writing score.

b. Covariates appearing in the model are evaluated at the following values: reading score = 52.23, social studies score = 52.41.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results12.png>

Looking at the table called Type III Tests of Fixed Effects, we can see that all of the predictor variables are statistically significant except for **prog**. Notice that the test of **prog** has two degrees of freedom (because the variable **prog** has three levels). In the table called Estimates of Fixed Effects we find the coefficients, and, as expected, the coefficient for the highest-numbered category of **female** and **prog** are 0 (because they are the reference groups). At the bottom of the output, there are two tables of Estimated Marginal Means. Looking at the table for **prog**, we can see that the estimated marginal mean for each level of **prog** is approximately 52. This fact can help us (and audience members) understand why **prog** is not a statistically significant predictor.

Random slopes

Another way that this model can be extended is by including a random slope. The inclusion of a random slope in a multilevel model should be done only when theory indicates the random slope is necessary or when testing a specific hypothesis. The random intercept is almost always included in a multilevel model, as it accounts for the clustering in the data; the random slope, on the other hand, is often optional. When a random slope is included, the random intercept and random slope might covary, in which case a covariance structure should be specified. If a covariance structure is not specified using the **covtype** option on the **random** subcommand, SPSS will use variance components (VC) by default. In the next example, the variable **socst** will be included on the **random** subcommand and an unstructured variance structure (UN) will be specified. The specification of the **covtype** option can be shortened to just **cov**.

```
mixed write with read socst by female prog
/fix read socst female prog
/random intercept socst | subject(cid) cov(un)
/print = solution.
```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	50.610	161.197	.000
read	1	189.440	10.876	.001
socst	1	20.070	7.650	.012
female	1	171.745	29.297	.000
prog	2	169.982	.196	.822

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	57.915632	4.384622	49.185	13.209	.000	49.105241	66.726023
read	-.211415	.064107	189.440	-3.298	.001	-.337870	-.084959
socst	.145041	.052440	20.070	2.766	.012	.035678	.254404
[female=0]	-3.721115	.687477	171.745	-5.413	.000	-5.078107	-2.364123
[female=1]	0 ^b	0
[prog=1]	-.246131	.996524	167.752	-.247	.805	-2.213475	1.721213
[prog=2]	.312534	.946496	170.278	.330	.742	-1.555844	2.180911
[prog=3]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter		Estimate	Std. Error
Residual		19.997638	2.239790
Intercept + socst [subject = cid]	UN (1,1)	118.814744	66.551514
	UN (2,1)	-.899123	.959790
	UN (2,2)	.017436	.016486

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results13.png>

The tables in the Fixed Effects part of the output look as we expect; the new information is in the table called Estimates of Covariance Parameters at the bottom of the output. The first line in the table gives the estimate and standard error of the level 1 residual. The second line gives the estimate of the variance and its standard error for the random intercept. The third line gives the covariance between the random intercept and the random slope and its standard error. We can see that this estimate is negative, meaning that higher the random intercept, the less steep the slope is. Conversely, the lower the intercept, the more steep the slope is. In the last line, the estimate of the variance around the random slope is given, along with its standard error. Clearly, there is much more variance around the grand intercept than around the grand slope. While the value of the variance around the random slope does not need to be interpreted, if it is very close to 0, it may be better to omit the random slope from the model.

Watch for error messages

Random slopes sometimes cause estimation problems, and not all models with random slopes will converge, as we see in the next example. Be careful to look for error messages in the output file immediately after the echo of the syntax, as SPSS will often provide the results from the last iteration, even if there was a problem with the estimation.

```
mixed write with read socst by female prog
/fixed read socst female prog
/random intercept read | subject(cid) cov(un)
/print = solution.
```

```
mixed write with read socst by female prog
/fixed read socst female prog
/random intercept read | subject(cid) cov(un)
/print = solution.
```

Mixed Model Analysis

Warnings

Iteration was terminated but convergence has not been achieved. The MIXED procedure continues despite this warning. Subsequent results produced are based on the last iteration. Validity of the model fit is uncertain.

[. \(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results14.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results14.png)

Choosing between the repeated and random subcommands

In the example using repeated measures data, the **repeated** subcommand was used to specify the repeated measures. However, the **random** subcommand could also have been used. One of differences between the **repeated** and **random** subcommands is what covaries. When the **repeated** subcommand is used, the covariance structure specifies the covariance between the time points. When the **random** subcommand is used, the covariance is between the random intercept and the random slope or slopes.

Tests of random effects

It seems reasonable to ask if these random effects are statistically significant. In fact, SPSS does offer a **testcov** option on the **print** subcommand that will provide such a test. However, not all statisticians believe that this is a valid test. The problem with this type of test is that the null hypothesis (namely that the random effect equals 0) is on the boundary of the parameter space. Please see [Parameter Estimation and Inference in the Linear Mixed Model](https://www.sciencedirect.com/science/article/pii/S002437951100320X) (https://www.sciencedirect.com/science/article/pii/S002437951100320X) by N. N. Gumedze and T. T. Dunne (2011) for more information.

Choosing a covariance structure

When running a model that has both a random intercept and a random slope, there is a covariance between the random intercept and the random slope. The default in SPSS is to use variance components (VC). When the variance components structure is specified on a **random** subcommand, a scaled identity (ID) structure is assigned to each of the effects specified on the subcommand. A scaled identity matrix has 1s on the diagonal (for the variances), and 0s on the off-diagonal (for the covariances). For many research situations, this is unrealistic, so a different type of covariance structure may be specified. Below are some examples of commonly used covariance matrices.

Identity

ID *Identity.* This is a scaled identity matrix.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

[. \(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_ID.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_ID.png)

Compound symmetry

CS *Compound symmetry.* This structure has constant variance and constant covariance.

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

[. \(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_CS.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_CS.png)

First order autoregressive

AR1 *First-order autoregressive.* The constraint $|\rho| \leq 1$ is imposed for stationarity.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

[. \(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_AR1.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_AR1.png)

Heterogeneous first order autoregressive

ARH1 *Heterogeneous first-order autoregressive.* The constraint $|\rho_k| \leq 1$ is imposed for stationarity.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho & \sigma_3 \sigma_1 \rho^2 & \sigma_4 \sigma_1 \rho^3 \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_3 \sigma_2 \rho & \sigma_4 \sigma_2 \rho^2 \\ - & - & - & - \\ - & - & - & - \end{bmatrix}$$

$$\begin{bmatrix} \sigma_3 \sigma_1 \rho^- & \sigma_3 \sigma_2 \rho & \sigma_3 & \sigma_4 \sigma_3 \rho \\ \sigma_4 \sigma_1 \rho^3 & \sigma_4 \sigma_2 \rho^2 & \sigma_4 \sigma_3 \rho & \sigma_4^2 \end{bmatrix}$$

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_ARH1.png)

[content/uploads/2018/02/cov_ARH1.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_ARH1.png)

Unstructured

UN *Unstructured.* This is a completely general covariance matrix.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

https://stats.idre.ucla.edu/wp-content/uploads/2018/02/cov_UN.png

Chapters 7 and 8 of Applied Longitudinal Data Analysis by Singer and Willett (2003) provide excellent descriptions and explanations of the issues.

Comparing models with different covariance structures

In some research situations it is clear which covariance structure should be used. In other research situations, however, the researcher may wish to test models that use different covariance structures to determine which is preferable. A likelihood ratio test (LRT), also known as a likelihood ratio chi-square test, is sometimes used for this purpose. By default, the SPSS **mixed** command uses restricted maximum likelihood (REML). While a likelihood ratio chi-square can be calculated with REML in some circumstances, it is not always appropriate. We are not going into this issue here, but you can get more information on this issue here: [Full ML v. Restricted ML \(http://www.ssicentral.com/hlm/help6/faq/Full_ML_vs_Restricted_ML.pdf\)](http://www.ssicentral.com/hlm/help6/faq/Full_ML_vs_Restricted_ML.pdf). For the example below, we will include the **method** subcommand with the **ml** option to change from the default to maximum likelihood.

```
* -2 log likelihood = 1242.431, number of parameters = 9.
mixed write with read socst by female prog
/fixd read socst female prog
/random intercept socst | subject(cid) cov(vc)
/method = ml
/print = solution.
```

Mixed Model Analysis

Model Dimension^a

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	read	1		1	
	socst	1		1	
	female	2		1	
	prog	3		2	
Random Effects	Intercept + socst ^b	2	Variance Components	2	cid
Residual				1	
Total		10		9	

a. Dependent Variable: writing score.

b. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

Information Criteria^a

-2 Log Likelihood	1242.431
Akaike's Information Criterion (AIC)	1260.431
Hurvich and Tsai's Criterion (AICC)	1261.379
Bozdogan's Criterion (CAIC)	1299.116
Schwarz's Bayesian Criterion (BIC)	1290.116

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results1>

```
* -2 log likelihood = 1241.234, number of parameters = 10.
mixed write with read socst by female prog
/fixd read socst female prog
/random intercept socst | subject(cid) cov(un)
```

```
/method = ml
/print = solution.
```

Mixed Model Analysis

Model Dimension^a

		Number of Levels	Covariance Structure	Number of Parameters	Subject Variables
Fixed Effects	Intercept	1		1	
	read	1		1	
	socst	1		1	
	female	2		1	
	prog	3		2	
Random Effects	Intercept + socst ^b	2	Unstructured	3	cid
Residual				1	
Total		10		10	

a. Dependent Variable: writing score.

b. As of version 11.5, the syntax rules for the RANDOM subcommand have changed. Your command syntax may yield results that differ from those produced by prior versions. If you are using version 11 syntax, please consult the current syntax reference guide for more information.

Information Criteria^a

-2 Log Likelihood	1241.234
Akaike's Information Criterion (AIC)	1261.234
Hurvich and Tsai's Criterion (AICC)	1262.398
Bozdogan's Criterion (CAIC)	1304.217
Schwarz's Bayesian Criterion (BIC)	1294.217

The information criteria are displayed in smaller-is-better form.

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results16.png>

The likelihood ratio test is easily calculated by hand. To do this, note the -2 log likelihood and the number of parameters from each model. In this example, the difference in -2 log likelihoods is very small (approximately 1.2), and the difference in degrees of freedom is 1. The critical value for a chi-square test with one degree of freedom is 3.841, which is well above our observed value of 1.2; hence, we conclude that the difference between the models is not statistically significant. Given this, the model with the variance components covariance is to be preferred. It should be acknowledged that not all statisticians believe that this is the best way to determine which covariance structure is best for a given model. Chuck Kincald gives a description of various methods that can be used to choose between covariance structures ([Guidelines for Selecting the Covariance Structure in Mixed Model Analysis](http://www2.sas.com/proceedings/sugi30/198-30.pdf) (<http://www2.sas.com/proceedings/sugi30/198-30.pdf>))

As a final note, here is a webpage with useful information on when to use and when not to use an unstructured covariance structure: [The Unstructured Covariance Matrix: When It Does and Doesn't Work](https://www.theanalysisfactor.com/unstructured-covariance-matrix-when-it-does-and-doesn%E2%80%99t-work/) (<https://www.theanalysisfactor.com/unstructured-covariance-matrix-when-it-does-and-doesn%E2%80%99t-work/>). The likelihood ratio chi-square works only when models are nested. In the example above, the models are nested because we are comparing a model using variance components to a model with an unstructured covariance structure. This test would not be valid if we compared the model with variance components to a model using AR1. When comparing models that are not nested, values such as AICC or BIC are often examined. With these types of measures, the model with the smaller value is to be preferred. However, caution should be used so that the model is not overfit to the data. A good discussion of the different types of information criteria and their limitations can be found in [The Comparison of Model Selection Criteria When Selecting Among Competing Hierarchical Linear Models](https://pdfs.semanticscholar.org/8b74/fb3018a6d65d9fdb63c1326f4fa311ed35f3.pdf) (<https://pdfs.semanticscholar.org/8b74/fb3018a6d65d9fdb63c1326f4fa311ed35f3.pdf>). Keselman, Algina, Kowalchuk and Wolfinger ([A Comparison of Two Approaches For Selecting Covariance Structures in The Analysis of Repeated Measurements](https://pdfs.semanticscholar.org/1d6c/f16816f05b6a80fa46611fdea22b680f596c.pdf) (<https://pdfs.semanticscholar.org/1d6c/f16816f05b6a80fa46611fdea22b680f596c.pdf>)) describe the limitations of using AIC or BIC when selecting models with different covariance structures.

ML v. REML

Even if you are not comparing models, you may still wonder whether to use ML or REML. REML is often preferred when the sample size is small (Raudenbush and Bryk, 2002). With these types of measures, the model with the smaller value is to be preferred. However, caution should be used so that the model is not overfit to the data. However, many sources that make this recommendation do not clearly state what is meant by "small". The default in SPSS is to use REML, but this can be changed by using the **ML** option on the **method** subcommand.

Interactions

Now let's talk about interactions. Multilevel models can include interaction terms, just as any other regression model can. The terms included in the interaction may be at level 1 or level 2 (or any higher level if the multilevel model has more than two levels). The terms in the interaction do not have to be at the same level. When the terms are not at the same level, the interaction is called a cross-level interaction. The example dataset does not have a level 2 predictor variable, so

one will be created for the purpose of the following example analysis. The first **sort cases** command is necessary for the **aggregate** command to work correctly; the second is simply to put the data back in the order they were before the new variable was created – it is technically not necessary.

```
sort cases by cid honors.
```



```
aggregate
/break = cid
/sum_honors = sum(honors).

sort cases by cid id.
exe.
```

Let's get some descriptive statistics on the new variable, `sum_honors`.

```
freq var = sum_honors.
```

Frequencies

Statistics		
sum_honors		
N	Valid	200
	Missing	0

sum_honors					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	81	40.5	40.5	40.5
	1.00	31	15.5	15.5	56.0
	2.00	10	5.0	5.0	61.0
	3.00	19	9.5	9.5	70.5
	4.00	17	8.5	8.5	79.0
	7.00	11	5.5	5.5	84.5
	8.00	10	5.0	5.0	89.5
	9.00	10	5.0	5.0	94.5
	10.00	11	5.5	5.5	100.0
	Total	200	100.0	100.0	

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results17.png>

```
means var = sum_honors.
```

Means

Case Processing Summary						
	Included		Cases Excluded		Total	
	N	Percent	N	Percent	N	Percent
sum_honors	200	100.0%	0	0.0%	200	100.0%

Report			
sum_honors			
Mean	N	Std. Deviation	
2.6650	200	3.31174	

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results18.png>

The variable `sum_honors` is ordinal, so it could be used as either a categorical or a continuous predictor variable, depending on our research purpose.

In our first example, we include an interaction of two level 1 variables. Because both of the variables are categorical, they are both listed after the SPSS keyword `by` on the `mixed` command. The specification of the interaction term is given on `fixed` subcommand.

```
mixed write by female prog
/fixed female prog female*prog
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female*prog).
```

Fixed Effects

Type III Tests of Fixed Effects^a

	Denominator	
--	-------------	--

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.853	887.617	.000
female	1	175.922	37.641	.000
prog	2	178.113	.695	.501
female * prog	2	175.954	2.479	.087

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
Intercept	55.058820	1.980228	29.662	27.804	.000	51.012718	59.104922
[female=0]	-6.009942	1.425429	175.848	-4.216	.000	-8.823092	-3.196791
[female=1]	0 ^b	0
[prog=1]	.158374	1.384053	175.372	.114	.909	-2.573169	2.889918
[prog=2]	-.607748	1.220237	177.504	-.498	.619	-3.015787	1.800290
[prog=3]	0 ^b	0
[female=0] * [prog=1]	.484136	2.047032	175.640	.237	.813	-3.555808	4.524080
[female=0] * [prog=2]	3.430189	1.737942	175.966	1.974	.050	.000296	6.860083
[female=0] * [prog=3]	0 ^b	0
[female=1] * [prog=1]	0 ^b	0
[female=1] * [prog=2]	0 ^b	0
[female=1] * [prog=3]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	23.149620	2.477631
Intercept [subject= cid] Variance	59.374788	20.271940

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results19.png)

[content/uploads/2018/02/results19.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results19.png)

Estimated Marginal Means

female * type of program^a

female	type of program	Mean	Std. Error	df	95% Confidence Interval	
					Lower Bound	Upper Bound
male	general	49.691	2.039	33.205	45.544	53.839
	academic	51.871	1.881	24.269	47.992	55.751
	vocation	49.049	2.052	33.884	44.877	53.220
female	general	55.217	2.006	31.203	51.126	59.308
	academic	54.451	1.852	22.845	50.618	58.284
	vocation	55.059	1.980	29.662	51.013	59.105

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results20.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results20.png)

Looking at the table called the Type III Tests of Fixed Effects, we can see that the predictor variable **female** is statistically significant, while **prog** and the **female** by **prog** interaction are not. Looking at the table called the Estimates of Fixed Effects, we see that neither of the coefficients for the variable **prog** are statistically significant, so this is consistent with the output in the table above. Looking at the interaction term, the coefficient for the **female** = 0 by **prog** = 1 term is not statistically significant, but the coefficient for female = 0 by **prog** = 2 has a p-value of 0.05, and the lower bound of the confidence interval does not include 0 (although it is very close). While some researchers might try to interpret this coefficient, most would not because the overall test of the interaction was not statistically significant.

Including a level 2 predictor

The next example does not contain an interaction term, but it does contain a level 2 predictor, **sum_honors**. Notice that nothing in the syntax indicates that the variable **sum_honors** is a level 2 variable.

```
mixed write by female sum_honors
/fix female sum_honors
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female)
/emmeans = tables(sum_honors).
```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Sum of Squares	df	Mean Square	F	Sig.
--------	----------------	----	-------------	---	------

Source	Numerator df	df	F	Sig.
Intercept	1	11.043	2705.645	.000
female	1	182.263	33.912	.000
sum_honors	8	11.098	8.610	.001

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	64.071336	3.915560	10.862	16.363	.000	55.439914	72.702758
[female=0]	-4.196174	.720573	182.263	-5.823	.000	-5.617912	-2.774437
[female=1]	0 ^b	0
[sum_honors=.00]	-17.171803	4.147406	10.804	-4.140	.002	-26.320456	-8.023151
[sum_honors=1.00]	-10.778490	4.517243	10.823	-2.386	.036	-20.740769	-8.816210
[sum_honors=2.00]	-6.314397	5.553037	10.986	-1.137	.280	-18.538502	5.909709
[sum_honors=3.00]	-3.737364	4.802643	10.927	-.778	.453	-14.316512	6.841785
[sum_honors=4.00]	-1.775898	4.821334	11.092	-.368	.720	-12.376868	8.825073
[sum_honors=7.00]	-.037407	5.528520	10.793	-.007	.995	-12.234089	12.159275
[sum_honors=8.00]	.687516	5.544318	10.917	.124	.904	-11.526713	12.901746
[sum_honors=9.00]	1.126751	5.544999	10.923	.203	.843	-11.088248	13.341750
[sum_honors=10.00]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	23.259055	2.457588
Intercept [subject= cid] Variance	13.148495	6.582433

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results21.png>

Estimated Marginal Means

1. female^a

				95% Confidence Interval	
female	Mean	Std. Error	df	Lower Bound	Upper Bound
male	55.653	1.174	13.768	53.130	58.175
female	59.849	1.160	13.144	57.345	62.353

a. Dependent Variable: writing score.

2. sum_honors^a

				95% Confidence Interval	
sum_honors	Mean	Std. Error	df	Lower Bound	Upper Bound
.00	44.801	1.391	11.053	41.743	47.860
1.00	51.195	2.269	11.020	46.202	56.188
2.00	55.659	3.940	11.133	47.000	64.317
3.00	58.236	2.796	11.299	52.101	64.370
4.00	60.197	2.825	11.747	54.028	66.367
7.00	61.936	3.908	10.780	53.313	70.559
8.00	62.661	3.936	11.096	54.006	71.316
9.00	63.100	3.934	11.067	54.448	71.752
10.00	61.973	3.908	10.780	53.350	70.596

a. Dependent Variable: writing score.

Before we get to the output, let's review the syntax that was used. Notice that **sum_honors**, a level 2 predictor, is specified in exactly the same way as **female**, a level 1 predictor.

Looking at the table called the Type III Tests of Fixed Effects, we can see that both **female** and **sum_honors** are statistically significant. Looking at the table called Estimates of Fixed Effects, we can see that two of the levels of **sum_honors** are statistically significant from the reference group while the others are not. Looking at the Estimated Marginal Means table for **sum_honors**, we can see why this is. The reference group for **sum_honors** is 10, with an estimated marginal mean of 61.973. The estimated marginal means for **sum_honors** = 0 and **sum_honors** = 1 have the greatest difference from 61.973. The values in this table do not necessarily need to be presented in a write up of the results. Rather, their purpose is to aid interpretation. They can be reported only if they help the audience of the research more clearly understand the results.

Interaction of a level 1 and level 2 predictor

Now let's interact the level 2 predictor variable **sum_honors** with a level 1 predictor variable, **female**.

```
mixed write by female sum_honors
```

```

/fixed female sum_honors female*sum_honors
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female*sum_honors).

```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	11.519	2763.726	.000
female	1	171.837	11.332	.001
sum_honors	8	11.464	9.300	.001
female * sum_honors	8	172.334	1.707	.100

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	61.571429	3.985606	12.501	15.448	.000	52.925969	70.216888
[female=0]	2.678571	2.979479	171.161	.899	.370	-3.202684	8.559827
[female=1]	0 ^b	0
[sum_honors=.00]	-13.970653	4.244347	12.695	-3.292	.006	-23.162414	-4.778892
[sum_honors=1.00]	-7.272457	4.616074	12.648	-1.575	.140	-17.273150	2.728237
[sum_honors=2.00]	-2.571429	6.311924	19.516	-.407	.688	-15.758815	10.615958
[sum_honors=3.00]	-2.177190	4.894659	12.636	-.445	.664	-12.782528	8.428148
[sum_honors=4.00]	-.746410	4.973871	13.465	-.150	.883	-11.454195	9.961376
[sum_honors=7.00]	2.678571	5.847324	14.461	.458	.654	-9.825249	15.182392
[sum_honors=8.00]	3.000000	5.636499	12.501	.532	.604	-9.226526	15.226526
[sum_honors=9.00]	2.428571	5.749901	13.530	.422	.679	-9.944010	14.801152
[sum_honors=10.00]	0 ^b	0
[female=0] * [sum_honors=.00]	-8.387244	3.180488	171.916	-2.637	.009	-14.665079	-2.109408
[female=0] * [sum_honors=1.00]	-9.474137	3.462837	171.376	-2.736	.007	-16.309441	-2.638833
[female=0] * [sum_honors=2.00]	-8.428571	4.795859	171.161	-1.757	.081	-17.895217	1.038075
[female=0] * [sum_honors=3.00]	-3.906143	3.793096	171.188	-1.030	.305	-11.393406	3.581119
[female=0] * [sum_honors=4.00]	-3.591175	3.830117	173.460	-.938	.350	-11.150809	3.968460
[female=0] * [sum_honors=7.00]	-7.214286	4.213620	171.161	-1.712	.089	-15.531637	1.103066
[female=0] * [sum_honors=8.00]	-6.250000	4.431439	171.161	-1.410	.160	-14.997310	2.497310

([https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results23.png)

[content/uploads/2018/02/results23.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results23.png))

[female=0] * [sum_honors=8.00]	-6.250000	4.431439	171.161	-1.410	.160	-14.997310	2.497310
[female=0] * [sum_honors=9.00]	-4.478571	4.232729	171.161	-1.058	.292	-12.833643	3.876501
[female=0] * [sum_honors=10.00]	0 ^b	0

[sum_honors=10.00]							
[female=1] * [sum_honors=.00]	0 ^b	0
[female=1] * [sum_honors=1.00]	0 ^b	0
[female=1] * [sum_honors=2.00]	0 ^b	0
[female=1] * [sum_honors=3.00]	0 ^b	0
[female=1] * [sum_honors=4.00]	0 ^b	0
[female=1] * [sum_honors=7.00]	0 ^b	0
[female=1] * [sum_honors=8.00]	0 ^b	0
[female=1] * [sum_honors=9.00]	0 ^b	0
[female=1] * [sum_honors=10.00]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	22.596754	2.442632
Intercept [subject= cid] Variance	12.656950	6.356882

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results24.png)

[content/uploads/2018/02/results24.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results24.png))

Estimated Marginal Means

female * sum_honors^a

female	sum_honors	Mean	Std. Error	df	95% Confidence Interval	
					Lower Bound	Upper Bound
male	.00	41.892	1.490	15.446	38.725	45.059
	1.00	47.503	2.475	16.654	42.273	52.734
	2.00	53.250	3.935	11.876	44.667	61.833
	3.00	58.167	3.177	20.031	51.540	64.794
	4.00	59.912	3.076	17.331	53.433	66.392
	7.00	59.714	3.986	12.501	51.069	68.360
	8.00	61.000	4.493	20.031	51.628	70.372
	9.00	62.200	4.144	14.597	53.345	71.055
	10.00	64.250	4.279	16.548	55.204	73.296
female	.00	47.601	1.459	14.298	44.477	50.724
	1.00	54.299	2.329	13.095	49.272	59.326
	2.00	59.000	4.894	27.764	48.970	69.030
	3.00	59.394	2.841	12.908	53.252	65.537
	4.00	60.825	2.976	15.489	54.500	67.150
	7.00	64.250	4.279	16.548	55.204	73.296
	8.00	64.571	3.986	12.501	55.926	73.217
	9.00	64.000	4.144	14.597	55.145	72.855
	10.00	61.571	3.986	12.501	52.926	70.217

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results25.png>)

Looking at the table called Type III Tests of Fixed Effects, we can see that the interaction of **female** and **sum_honors** is not statistically significant. Looking at the table called Estimates of Fixed Effects, we see that two coefficients in the interaction term are statistically significant; again, this would not be interpreted because the overall interaction term (the 8 df test) is not statistically significant. Comparing the results for the variable **female** in the table of Type III Tests of Fixed Effects and the Estimates of Fixed Effects, we see that in the former table, the p-value for **female** is 0.001, while in the latter table the p-value is 0.370. Why are these p-values so different? The difference is caused by the type of coding scheme used for the categorical variables. In the table of Type III Tests of Fixed Effects, effect coding is used for the categorical variable. However, in the table of Estimates of Fixed Effects, dummy coding is used. This difference matters only when there is an interaction term in the model and one or more of the variables in the interaction is categorical. The p-value of the interaction term is not affected, but the p-values of the lower-order terms may be. Oftentimes, the lower-order effects are not interpreted (although they are always reported), so the difference is of minimal consequence. It is not the case that one p-value is correct and the other is not; rather, they give different information. When reporting the results, it is important to be clear which is being reported. The results in the table Type III Tests of Fixed Effects are main effects, while the results in the table estimates of Fixed Effects are simple effects. Main effects are deviations from the grand mean, while simple effects are the effect at a specific value. For these

results, we would say that the effect of being male compared to female is 2.678571 at the reference level of **prog** and the reference level of the interaction term. This test may or may not be of interest to the researcher. While the coefficients and their p-values are always reported, the simple effect may or may not be interpreted.

The variable **sum_honors** is ordinal, so it is possible to treat it as continuous. In the following example, **sum_honors** is treated as a continuous predictor. On the

The variable `sum_honors` is ordinal, so it is possible to treat it as continuous. In the following example, `sum_honors` is treated as a continuous predictor. On the `emmeans` subcommand, it has been held at its mean, which is 2.665 (rounded to 2.67).

```
mixed write by female with sum_honors
/fixed female sum_honors female*sum_honors
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female) with (sum_honors = 2.67).
```

Estimated Marginal Means

female ^a					
female	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
male	50.657 ^b	1.083	23.595	48.419	52.895
female	54.850 ^b	1.060	21.708	52.650	57.050

a. Dependent Variable: writing score.

b. Covariates appearing in the model are evaluated at the following values: `sum_honors = 2.67`.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results27.png>

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	18.013	1311.962	.000
female	1	182.515	40.298	.000
sum_honors	1	17.930	43.480	.000
female * sum_honors	1	180.779	7.866	.006

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	50.162288	1.368148	21.683	36.664	.000	47.322512	53.002064
[female=0]	-5.774648	.909675	182.515	-6.348	.000	-7.569479	-3.979817
[female=1]	0 ^b	0
sum_honors	1.755839	.324208	21.086	5.416	.000	1.081780	2.429898
[female=0] * sum_honors	.592212	.211156	180.779	2.805	.006	.175565	1.008859
[female=1] * sum_honors	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	22.470487	2.381325
Intercept [subject = cid] Variance	18.169742	6.812788

a. Dependent Variable: writing score.

<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results26.png>

[content/uploads/2018/02/results26.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results26.png)

In this model, the interaction as well as both of the lower-order effects, is statistically significant.

Tests of simple main effects

When there is an interaction in the model, a test of simple main effects is sometimes useful. A test of simple main effects asks if the effect of one variable is statistically significant at a given level of another variable. This test can be obtained by using the `compare` option on the `emmeans` subcommand. Let's return to the earlier example that included the interaction of the two level 1 variables `female` and `prog`.

```
mixed write by female prog
/fixed female prog female*prog
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female*prog) compare(prog).
```

Fixed Effects

Type III Tests of Fixed Effects^a

	Denominator	
--	-------------	--

Source	Numerator df	df	F	Sig.
Intercept	1	18.853	887.617	.000
female	1	175.922	37.641	.000
prog	2	178.113	.695	.501
female * prog	2	175.954	2.479	.087

a. Dependent Variable: writing score.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	55.058820	1.980228	29.662	27.804	.000	51.012718	59.104922
[female=0]	-6.009942	1.425429	175.848	-4.216	.000	-8.823092	-3.196791
[female=1]	0 ^b	0
[prog=1]	.158374	1.384053	175.372	.114	.909	-2.573169	2.889918
[prog=2]	-.607748	1.220237	177.504	-.498	.619	-3.015787	1.800290
[prog=3]	0 ^b	0
[female=0] * [prog=1]	.484136	2.047032	175.640	.237	.813	-3.555808	4.524080
[female=0] * [prog=2]	3.430189	1.737942	175.966	1.974	.050	.000296	6.860083
[female=0] * [prog=3]	0 ^b	0
[female=1] * [prog=1]	0 ^b	0
[female=1] * [prog=2]	0 ^b	0
[female=1] * [prog=3]	0 ^b	0

a. Dependent Variable: writing score.

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

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	23.149620	2.477631
Intercept [subject = cid] Variance	59.374788	20.271940

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results28.png)

[content/uploads/2018/02/results28.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results28.png)

Estimated Marginal Means

female * type of program

Estimates^a

female	type of program	Mean	Std. Error	df	95% Confidence Interval	
					Lower Bound	Upper Bound
male	general	49.691	2.039	33.205	45.544	53.839
	academic	51.871	1.881	24.269	47.992	55.751
	vocation	49.049	2.052	33.884	44.877	53.220
female	general	55.217	2.006	31.203	51.126	59.308
	academic	54.451	1.852	22.845	50.618	58.284
	vocation	55.059	1.980	29.662	51.013	59.105

a. Dependent Variable: writing score.

Pairwise Comparisons^a

female	(I) type of program	(J) type of program	Mean Difference (I-J)	Std. Error	df	Sig. ^c	95% Confidence Interval for Difference ^c	
							Lower Bound	Upper Bound
male	general	academic	-2.180	1.328	176.191	.103	-4.801	.442
		vocation	.643	1.562	176.884	.681	-2.441	3.726
	academic	general	2.180	1.328	176.191	.103	-.442	4.801
		vocation	2.822 [*]	1.396	178.657	.045	.067	5.578
	vocation	general	-.643	1.562	176.884	.681	-3.726	2.441
		academic	-2.822 [*]	1.396	178.657	.045	-5.578	-.067
female	general	academic	.766	1.277	177.553	.549	-1.754	3.286
		vocation	.158	1.384	175.372	.909	-2.573	2.890
	academic	general	-.766	1.277	177.553	.549	-3.286	1.754
		vocation	-.608	1.220	177.504	.619	-3.016	1.800
	vocation	general	-.158	1.384	175.372	.909	-2.890	2.573
		academic	.608	1.220	177.504	.619	-1.800	3.016

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Dependent Variable: writing score.

c. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results29.png)[content/uploads/2018/02/results29.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results29.png)**Pairwise Comparisons^a**

female	(I) type of program	(J) type of program	Mean Difference (I-J)	Std. Error	df	Sig. ^c	95% Confidence Interval for Difference ^c	
							Lower Bound	Upper Bound
male	general	academic	-2.180	1.328	176.191	.103	-4.801	.442
		vocation	.643	1.562	176.884	.681	-2.441	3.726
	academic	general	2.180	1.328	176.191	.103	-.442	4.801
		vocation	2.822 [*]	1.396	178.657	.045	.067	5.578
	vocation	general	-.643	1.562	176.884	.681	-3.726	2.441
		academic	-2.822 [*]	1.396	178.657	.045	-5.578	-.067
female	general	academic	.766	1.277	177.553	.549	-1.754	3.286
		vocation	.158	1.384	175.372	.909	-2.573	2.890
	academic	general	-.766	1.277	177.553	.549	-3.286	1.754
		vocation	-.608	1.220	177.504	.619	-3.016	1.800
	vocation	general	-.158	1.384	175.372	.909	-2.890	2.573
		academic	.608	1.220	177.504	.619	-1.800	3.016

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Dependent Variable: writing score.

c. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests^a

female	Numerator df	Denominator df	F	Sig.
male	2	177.251	2.552	.081
female	2	176.922	.222	.801

Each F tests the simple effects of type of program within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results30.png)[content/uploads/2018/02/results30.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results30.png)

Be aware that not all statisticians approve of the use of tests of simple main effects. In particular, there are concerns over the conceptual error rate. Tests of simple main effects are one tool that can be useful in interpreting interactions. Caution should be exercised in interpreting the results of analyses of simple main effects. In general, the results of tests of simple main effects should be considered suggestive and not definitive.

Amount of variance explained

Amount of variance explained

As was made clear earlier in this workshop, the SPSS **mixed** command is used to run linear models, models that are, in many ways, similar to OLS regression. The R-squared value associated with an OLS model is often used to describe the amount of variance in the outcome variable that is explained by the predictor variables. Unfortunately, there is no obvious equivalent in a multilevel model. However, there are some approximations that can be calculated. There are good discussions of this topic Snijders and Bosker (2012), chapter 7 and Singer and Willett (2003) pages 102-104.

Testing custom hypotheses

Let's look at a few examples of the **test** subcommand. The **test** subcommand is used to test custom hypotheses. Specifically, you specify the null hypothesis as a linear combinations of parameters. You can use multiple **test** subcommands in a given call to **mixed** and each is treated independently. Multiple linear combinations can also be specified on a single **test** subcommand, and when this is done, they need to be separated with a semicolon. The contrast coefficients listed on the **test** subcommand should always sum to 0. In our first example, we will use effect coding to test the effect of the variable **female**. As a side note, the code below is actually the multilevel version of a t-test.

```
mixed write by female
/fixd female
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female)
/test = 'using effect coding' female 1 -1.
```

Custom Hypothesis Test (using effect coding)

Contrast Estimates^{a,b}

Contrast	Estimate	Std. Error	df	Test Value	t	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
L1	-4.138289	.722593	180.509	0	-5.727	.000	-5.564106	-2.712473

a. using effect coding

b. Dependent Variable: writing score.

Estimated Marginal Means

female^a

female	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
male	50.644	1.847	20.906	46.802	54.485
female	54.782	1.832	20.272	50.963	58.601

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results31.png)

[content/uploads/2018/02/results31.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results31.png))

We can see the results in the section of the output called Custom Hypothesis Test in the table called Contrast Estimates. The contrast estimate is -4.138. The output from the **emmeans** subcommand shows the estimated marginal means for each level of **female**. Because we used effect coding, the contrast coefficient is the difference between these two means.

Interaction contrasts can be done using the **test** subcommand. Let's say that we want to compare the average program type 1 and 2 versus 3 and female 0 versus 1. Before trying to write the syntax, let's make a table with the two categorical predictor variables, **female** and **prog**, along with the values of the contrast coefficients.

prog 1 = -.5 prog 2 = -.5 prog 3 = 1

female 0 =	1	-.5	-.5	1
female 1 =	-1	.5	.5	-1

```
mixed write by female prog
/fixd female prog female*prog
/random intercept | subject(cid)
/print = solution
/emmeans = tables(female*prog)
/test = 'prog12v3 & female 0v1' prog*female -.5 -.5 1 .5 .5 -1.
```

Custom Hypothesis Test (prog12v3 & female 0v1)

Contrast Estimates^{a,b}

Contrast	Estimate	Std. Error	df	Test Value	t	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
L1	-1.957163	1.669946	175.664	0	-1.172	.243	-5.252903	1.338577

a. prog12v3 & female 0v1

b. Dependent Variable: writing score.

Estimated Marginal Means

female * type of program^a

female	type of program	Mean	Std. Error	df	95% Confidence Interval	
					Lower Bound	Upper Bound
male	general	49.691	2.039	33.205	45.544	53.839
	academic	51.871	1.881	24.269	47.992	55.751
	vocation	49.049	2.052	33.884	44.877	53.220
female	general	55.217	2.006	31.203	51.126	59.308
	academic	54.451	1.852	22.845	50.618	58.284
	vocation	55.059	1.980	29.662	51.013	59.105

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results32.png)[content/uploads/2018/02/results32.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results32.png)

We can see the results in the table called Contrast Estimates. None of the three contrasts that were specified are statistically significant.

In the next example, multiple linear combinations will be specified on a single **test** subcommand. The variable **prog**, which has three levels, will be used. The first will compare level 1 of **prog** with level 3; the second will compare level 1 with level 2; and the third will compare level 2 with level 3. All of these comparisons will be divided by 3. This is done only to show an example of how to use the **divisor** option.

```
mixed write by prog
/fixd prog
/random intercept | subject(cid)
/print = solution
/emmeans = tables(prog)
/test = 'contrasts of prog' prog 1 0 -1; prog 1 -1 0; prog 0 1 -1 divisor = 3.
```

Custom Hypothesis Test (contrasts of prog)

Contrast Estimates^{a,b}

Contrast	Estimate	Std. Error	df	Test Value	t	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
L1	.396136	1.150412	179.791	0	.344	.731	-1.873911	2.666182
L2	-.644128	1.023240	180.988	0	-.629	.530	-2.663142	1.374886
L3	.346754	.353296	183.347	0	.981	.328	-.350294	1.043803

a. contrasts of prog

b. Dependent Variable: writing score.

Test of Contrasts^a

Source	Numerator df	Denominator df	F	Sig.
contrasts of prog	2	181.423	.517	.597

a. Dependent Variable: writing score.

Estimated Marginal Means

type of program^a

type of program	Mean	Std. Error	df	95% Confidence Interval	
				Lower Bound	Upper Bound
general	52.677	1.930	25.178	48.704	56.650
academic	53.321	1.832	20.497	49.506	57.136
vocation	52.281	1.933	25.281	48.302	56.260

a. Dependent Variable: writing score.

[https://stats.idre.ucla.edu/wp-](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results33.png)[content/uploads/2018/02/results33.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results33.png)

The 1 degree of freedom tests are shown in the table Contrast Estimates; the 2 degree of freedom test is given in the table called Tests of Contrasts. We can see that none of these tests are statistically significant.

Saving variables to the active dataset

The **save** subcommand can be used to save any of seven different variables related to the current multilevel model to the active dataset. To do this, the temporary variable name must be specified on the **save** subcommand, and an optional new name can be specified in parentheses. In the following example, five new variables will be saved to the dataset. These are the fixed predicted values, which are the regression means without the random effects

(called **blue**); the predicted values, which are the model fitted values (called **blup**); the standard error of the fixed predicted values (called **seblue**); the standard errors of the predicted values (called **seblup**); and the level 1 residuals (called **resid1**).

```
mixed write with read socst by female prog
/fixd read socst female prog
/random intercept socst | subject(cid) cov(un)
/print = solution
/save = fixpred(blue) pred(blup) sefixp(seblue) sepred(seblup) resid (resid1).
```

```
list cid id write blue blup seblue seblup resid1
/cases from 1 to 10.
```

```
cid id write blue blup seblue seblup resid1
```

```
1 1 44 56.67 40.42 2.35 1.47 3.58
1 2 41 55.62 39.37 2.23 1.46 1.63
1 15 39 52.04 35.80 2.26 1.44 3.20
1 45 35 54.50 38.11 2.58 1.96 -3.11
1 51 36 54.45 38.18 2.22 1.64 -2.18
1 53 37 51.50 35.16 2.51 1.60 1.84
1 67 37 51.01 34.68 2.42 1.53 2.32
1 108 33 51.98 35.68 2.46 1.62 -2.68
1 133 31 48.12 31.77 2.30 1.77 -.77
1 153 31 53.35 37.19 2.24 1.99 -6.19
```

Number of cases read: 10 Number of cases listed: 10

The acronym BLUP stands for best linear unbiased prediction, and it is used in linear mixed models for the estimation of random effect. It is similar to the best linear unbiased estimates (BLUEs) of fixed effects. Please see [That BLUP is a good thing: The estimation of random effects](https://projecteuclid.org/download/pdf_1/euclid.ss/1177011926) (https://projecteuclid.org/download/pdf_1/euclid.ss/1177011926) for more information.

If you change the model but not the names of the variables to be saved to the dataset, you will get an error message indicating that the variable names given on the **save** subcommand name variables already in the dataset. You can use the **delete variables** command to remove the variables from the dataset.

```
delete variables blue blup seblue seblup resid1.
```

```
mixed write with read by female
/fixd read socst female prog
/random intercept socst | subject(cid) cov(un)
/print = solution
/save = fixpred(blue) pred(blup) sefixp(seblue) sepred(seblup) resid(resid1).
```

Checking model assumptions

For the most part, the assumptions of OLS regression also apply to linear multilevel models. However, most of the assumptions apply to each level of the multilevel model.

- 1. Independence of observations** This assumption applies only at the highest level of nesting. Also, it is assumed that no important level of nesting is omitted from the model. If a level of nesting is omitted, the variance associated with that level will be mainly distributed to the level immediately below, as well as the level immediately above, if there is a level above (Snijder and Bosker, 2012).
- 2. Correct specification of the model** This means that all of the necessary variables are specified for both the fixed and random parts of the model for each level of the model. It also means that no unnecessary variables are included.
- 3. Functional form** The predictors are linearly related to the outcome. To test this assumption at level 1, a graph would be made for each level 2 unit of the outcome and each predictor. To test this assumption at level 2, plots of OLS estimates against each level 2 predictor would be made.
- 4. Homoscedascity** The residuals have a constant variance. Heteroscedascity may be observed if an important variable is omitted from the model. To test this assumption, plots of the raw residuals against each predictor would be made.
- 5. Normality of residuals** All residuals in the model are normally distributed. Many researchers, including Singer and Willett (2003, page 128) prefer visual inspection of such graphs rather than formal tests such as Wilks-Shapiro or Kolmogorov-Smirnov.

Checking the assumptions of even a simple multilevel model is a lot of work. For the first two assumptions, it is up to the researcher to know these things about the data and the model. The other two assumptions can be assessed with graphs. We will take a very simple model as an example. The **save** subcommand will be used to save the necessary variables to the dataset.

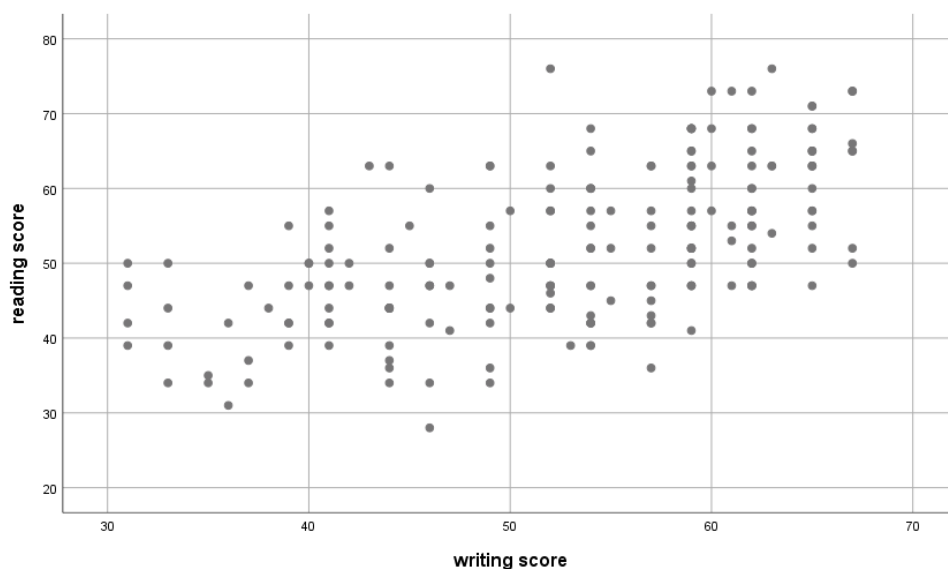
```
* checking model assumptions.
mixed write with read
/fixd read
/random intercept | subject(cid)
```

```
/print = solution
/save = fixpred(blue_read) pred(blup_read) resid(resid1_read).
```

** Output omitted.

```
* linearity.
graph
/scatterplot write with read.
```

Graph

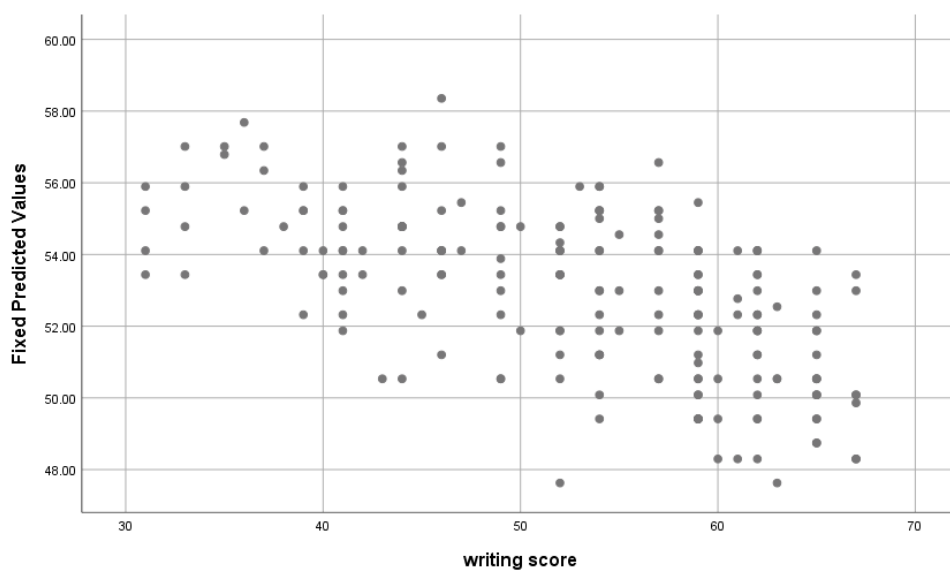


<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results34.png>

The relationship between **write** and **read** appears to be mostly linear.

```
* homogeneity.
graph
/scatterplot write with blue_read.
```

Graph



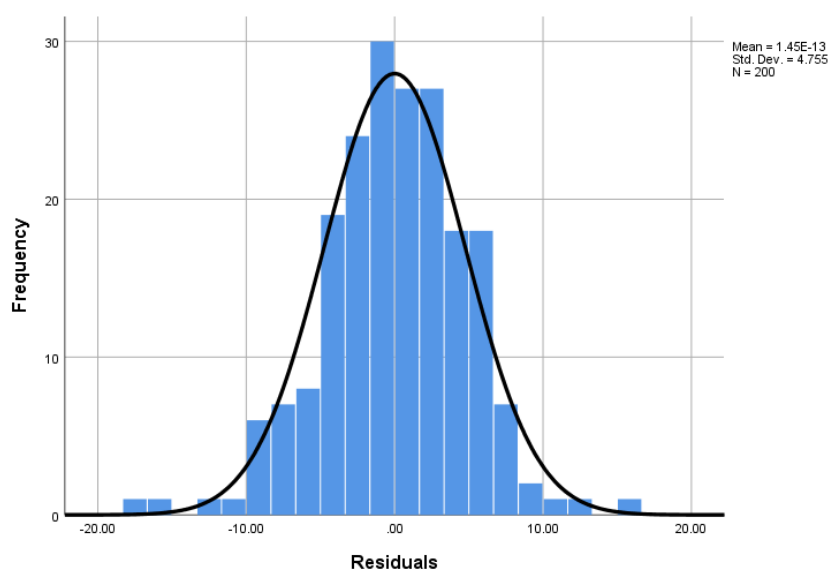
<https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results35.png>

The variance seems to be smaller at lower values of **write**.

```
* normality.
graph
/scatterplot (resid1_read) with (blue_read)
```

```
/nistogram(normal) = resid1_read.
```

Graph



[content/uploads/2018/02/results36.png](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results36.png)

The residuals appear to be approximately normally distributed.

An example of a three-level model

Let's look at an example of a three-level model. In this example, we have students nested within classrooms nested within schools. Notice that two **random** subcommands are needed. The first specifies the highest level of the model, and the second specifies the second-highest level nested in the highest level. This example can be expanded to run models with four, five or even more levels. The dataset [eg_hlm_small](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/eg_hlm_small.sav) (https://stats.idre.ucla.edu/wp-content/uploads/2018/02/eg_hlm_small.sav) is used here.

```
mixed math with year
/fixd = year
/print = solution
/random = intercept | subject(school)
```

```
/random = intercept | subject(school*cid).
```

Fixed Effects

Type III Tests of Fixed Effects^a

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	13.154	29.848	.000
year	1	1406.872	4370.084	.000

a. Dependent Variable: math test in an IRT scale score metric.

Estimates of Fixed Effects^a

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	-.669488	.122541	13.154	-5.463	.000	-.933909	-.405067
year	.752005	.011376	1406.872	66.107	.000	.729690	.774320

a. Dependent Variable: math test in an IRT scale score metric.

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	.385631	.014869
Intercept [subject = school]	Variance .178546	.081510
Intercept [subject = school * cid]	Variance .627458	.051385

a. Dependent Variable: math test in an IRT scale score metric.

[_\(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results3](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results3)

Other topics

For this last section, let's discuss some common issues regarding multilevel models.

Intraclass correlation

As we know, the observations within clusters are not independent. We can assess the extent of the non-independence of observations within clusters by calculating the intraclass correlation. To do this, we run a null model (i.e., a model with no predictor variables, which can also be called an empty model, and intercept-only model or an unconditional means model) and then do a little simple math by hand. We will use the hsbdemo dataset that we used before for this example. First, the formula:

$$\frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2}$$

[_\(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/ICC.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/ICC.png)

The numerator is the between-cluster variance, and the denominator is the sum of the between-cluster and within-cluster variance. The SPSS syntax needed to get these two numbers is minimal.

```
mixed write
```

```
/random intercept | subject(cid).
```

Covariance Parameters

Estimates of Covariance Parameters^a

Parameter	Estimate	Std. Error
Residual	27.297252	2.877135
Intercept [subject = cid]	Variance 64.360990	21.762601

a. Dependent Variable: writing score.

[_\(https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results38.png\)](https://stats.idre.ucla.edu/wp-content/uploads/2018/02/results38.png)

Looking at the table called Estimates of Covariance Parameters at the bottom of the output, we can see that the between-cluster variance is 64.361 and the within-cluster variance is 27.297. Hence, the ICC is 0.70. Values of ICC can range from 0 to 1, so a value of .7 is fairly large. Reporting an ICC value is often a requirement of peer-reviewed journals. Please note that there are at least 10 different definitions of ICC, so when reporting an ICC, you need to be clear about which version of ICC you are reporting. Please see [A guideline of selecting and reporting intraclass correlation coefficients for reliability research](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/>) for more information.

Design effect

Another reason to calculate an ICC is to determine if the non-independence in the data is "strong" enough to warrant the use of a multilevel model. If the data are nearly independent and all of the variables in the model are at the same level, a single-level model would be preferred because it is more parsimonious. To determine if a multilevel model is really needed, the design effect can be calculated. Again, this is a simple formula with math that can be done by hand.

design effect = $1 + (\text{average_cluster_size} - 1) * \text{ICC}$

The ICC can be calculated as shown above. The average cluster size can be found with the use of the **aggregate** command.

```
aggregate outfile = "D:\data\temp\mixed_n.sav"
/break = cid
/num_cid = nu(cid).

get file "D:\data\temp\mixed_n.sav".
mean num_cid.
```

Means

D:\data\temp\mixed_n.sav

Case Processing Summary

	Included		Cases Excluded		Total	
	N	Percent	N	Percent	N	Percent
num_cid	20	100.0%	0	0.0%	20	100.0%

Report

num_cid		
Mean	N	Std. Deviation
10.00	20	1.124

https://stats.idre.ucla.edu/wp-content/uploads/2019/03/means_output.png

We can see that the average cluster size is 10.00. So, design effect = $1 + (10 - 1) * .7 = 7.3$. If the design effect is less than 2 (Muthen and Satorra, 1995), we can safely ignore the clustering in the data. In this example, the design effect is much greater than 2, so a multilevel model (or some other method) is needed to account for the clustering in the data.

Denominator degrees of freedom

There is more than one way to calculate the denominator degrees of freedom in a multilevel model. SPSS uses the Satterthwaite approximation to calculate the denominator degrees of freedom, and this method is valid for both balanced and unbalanced designs. Other statistical software packages use other methods, so the results of the fixed effects may not match up between SPSS and SAS, for example. Please see [IBM Support page \(http://www-01.ibm.com/support/docview.wss?uid=swg21477296\)](http://www-01.ibm.com/support/docview.wss?uid=swg21477296) for more information.

Centering

Centering is a big topic that will not be covered in detail in this workshop. For those not familiar with the term centering, “centering” means subtracting a mean from each value of a variable. Level 1 variables can be centered around their group means or the grand mean; level 2 variables can be grand-mean centered. In times past, centering was almost expected. Currently, it is suggested that centering be used with caution and good reason. There are references at the end of with workshop that provide further information. Please note that statistical thinking regarding centering has evolved over time, so it may be important to note the publication date of references. Please see [HLM centering \(http://www.ssicentral.com/hlm/techdocs/Centering.pdf\)](http://www.ssicentral.com/hlm/techdocs/Centering.pdf) and [Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects \(https://www.ncbi.nlm.nih.gov/pubmed/16394187\)](https://www.ncbi.nlm.nih.gov/pubmed/16394187) for more information.

Sample size

The final topic is that sample size, and one of the most common questions regarding sample size is “how small is too small?”. One widely-cited article by Cora Mass and Joop Hox from 2005 has some specific recommendations based on simulations that they ran. The short summary of their findings is that about 50 level 2 observations are needed to avoid bias in the parameter estimates. Since then, many have questioned the need for 50 level 2 units, and indeed, much research using multilevel models has been published with fewer than 50 level 2 units. It is worth remembering, however, that multilevel models (at least those shown here) run on maximum likelihood, which is based on asymptotics, which necessarily requires a large sample size. Also, it is known that maximum likelihood can give biased results when run with small sample sizes; but the question of “how small is too small?” remains. Again, this is a big topic that is only being mentioned here; please see the references below for more information. If a power analysis for a multilevel model is needed, there are few off-the-shelf software packages available. One free option is: [ML PowerSim \(http://seis.bris.ac.uk/~frwjb/esrc.html#_ga=2.135410225.1875062554.151793.7472-176776403.1499977293\)](http://seis.bris.ac.uk/~frwjb/esrc.html#_ga=2.135410225.1875062554.151793.7472-176776403.1499977293). Another option is to use Optimal Design (<http://hlmsoft.net/od/>). Be aware that such power analyses usually require information that is not readily available and is not part of the standard multilevel output.

Imperfect hierarchies

In all of our previous examples, the hierarchy of observations has been very clear: each level 1 observation was clustered in one and only one level 2 unit. Not all data are so neatly structured. Imagine what our data would look like if some students started the academic year in one classroom, and then switched to another classroom at some point during the year. Now those students would be associated with more than one classroom. This is an example of a multiple membership model. With cross-classified data, each lower-level unit belongs to one higher-level unit (call it A), and also to another type of higher-level unit (call it B), but the higher-level units are not nested either way (meaning that A is not nested in B, and B is not nested in A). In multiple membership multiple classification models, both situations mentioned above are present. Chapter 13 in Snijders and Bosker (2012) gives a detailed explanation of these models, as well as examples of

how to analyze such data. We do not have an example dataset to illustrate the analysis of imperfect hierarchies. Some sample syntax for a cross-classified model, taken from [Multilevel Cross-Classified and Multi-Membership Models \(https://www.ihrp.uic.edu/files/CrossClassMultiMember-Hedeker2014.pdf\)](https://www.ihrp.uic.edu/files/CrossClassMultiMember-Hedeker2014.pdf) is shown below.

```
mixed attain with vrq
/fixd vrq
/print = solution
/random intercept | subject(pid)
/random intercept | subject(sid).
```

For more information on the SPSS mixed command

Please see the SPSS Command Syntax Reference for the most up-to-date information about the SPSS **mixed** command. You can access the Syntax Reference Guide by clicking on Help -> Command Syntax Reference.

References

- De Leeuw, J. Centering in linear multilevel models. In Encyclopedia of Statistics in Behavioral Science. Everitt, B. S. and Howell, D. C., editors. volume 1, pages 247-249. <http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa085/pdf>
- Gumedze, F. N. and Dunne, T. T. Paramter estimation and inference in the linear mixed model. *Linear Algebra and its Applications*, 435(8), 1920-1944. <https://www.sciencedirect.com/science/article/pii/S002437951100320X>
- Harrell, Jr. F. E. Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis, second edition. 2015. New York: Springer.
- Heck, R. H., Thomas, S. L. and Tabata, L. N. (2014). Multilevel and Longitudinal Modeling with IBM SPSS, second edition. New York: Routledge.
- Hedeker, D. Multilevel cross-classified and multi-membership models <https://www.ihrp.uic.edu/files/CrossClassMultiMember-Hedeker2014.pdf>
- http://www.unc.edu/~dbauer/manuscripts/BauerSterba_appendix_SPSS.pdf
- IBM Support: Denominator degrees of freedom for fixed effects in SPSS mixed
- http://www.unc.edu/~dbauer/manuscripts/BauerSterba_appendix_SPSS.pdf
- Kelly, J., Evans, M. D. R., Lowman, J. and Kykes, V. (2017). Group-mean-centering independent variables in multilevel models is dangerous. *Quality and Quantity*, 51(1), 261-283. <https://link.springer.com/article/10.1007/s11135-015-0304-z>
- Kenny, D. A. and Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, 99(3), 422-431. <http://psycnet.apa.org/record/1986-21031-001>
- Keselman, H. J., Algina, J., Kowalchuk, R. K. and Wolfinger, R. D. A comparison of two approaches for selecting covariance structures in the analysis of repeated measures. <https://pdfs.semanticscholar.org/1d6c/f16816f05b6a80fa46611dea22b680f596c.pdf>
- Kincaid, C. Guidelines for Selecting the covariance structure in mixed model analysis. <http://www2.sas.com/proceedings/sugi30/198-30.pdf>
- Koo, T. K. and Li, M. Y. (2016). A guideline for selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4913118/>
- Leyland, A. H. A review of multilevel modelling in SPSS (2004). <https://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/reviewsspss.pdf>
- Muthen, B. O. and Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, vol. 25, pages 267-316. <http://www.jstor.org/stable/271070>
- Paccagnella, O. Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. (2006). *Evaluation Review*, 30(1), 66-85. <https://www.ncbi.nlm.nih.gov/pubmed/16394187>
- Raudenbush, S. W. and Bryk, A. S. Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition. (2002). Thousand Oaks, CA: Sage Publications.
- Robinson, G. K. That BUP is a good thing: The estimation of random effects. (1991). *Statistical Science*, 6(1), 15-2. <https://projecteuclid.org/euclid.ss/1177011926>
- Scariano, S. M. and Davenport, J. M. The effects of violations of independence assumptions in the one-way ANOVA. (1987). *The American Statistician*, 41(2), 123-129. <http://www.jstor.org/stable/2684223>
- Singer, J. D. and Willett, J. B. Applied longitudinal data analysis: Modeling change and event occurrence. 2003. New York: Oxford University Press.
- Snijders, T. A. B. and Bosker, R. J. Multilevel analysis: An introduction to basic and advanced multilevel modeling, second edition. 2012. Thousand Oaks, CA: Sage Publications.
- Whittaker, T. A. and Furlow, Carolyn F. (2009). The Comparison of Model Selection Criteria When Selecting Among Competing Hierarchical Linear Models. *Journal of Modern Applied Statistical Methods*. 8(1), 173-193. <https://pdfs.semanticscholar.org/8b74/fb3018a6d65d9fdb63c1326f4fa311ed35f3.pdf>
- Full ML v. Restricted ML http://www.ssicentral.com/hlm/help6/faq/Full_ML_vs_Restricted_ML.pdf
- The Unstructured Covariance Matrix: When It Does and Doesn't Work <https://www.theanalysisfactor.com/unstructured-covariance-matrix-when-it-does-and-doesn%E2%80%99t-work/>

[Click here to report an error on this page or leave a comment](#)

[How to cite this page \(https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/\)](https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/)

