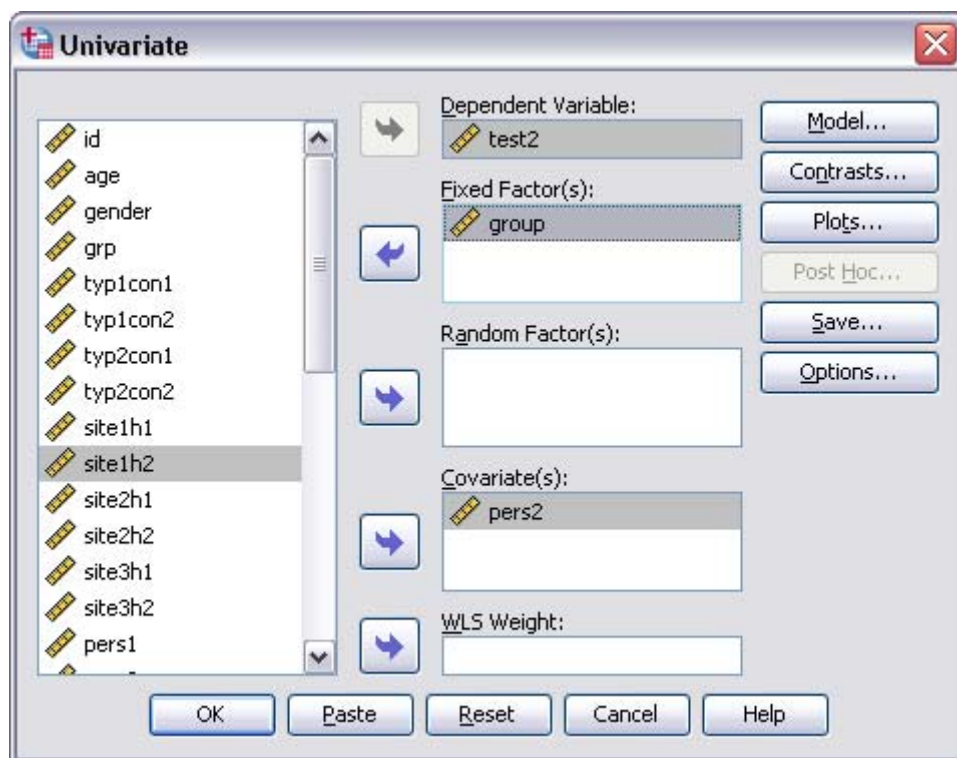


# Using the GLM Procedure in SPSS



Alan Taylor, Department of Psychology  
Macquarie University

2002-2011

© Macquarie University 2002-2011



<b><u>Introduction</u></b>	<b>1</b>
<b><u>1. General</u></b>	<b>3</b>
1.1 Factors and Covariates	3
1.1.1 Fixed Factors	4
1.1.2 Covariates	4
1.2 Sums of Squares	4
1.3 Interactions	5
1.4 Layout of Data	7
1.5 A Priori and Post Hoc Comparisons	8
1.6 Bonferroni Adjustment	9
1.7 'Protected' Post Hoc Tests	9
1.8 Orthogonal and Non-Orthogonal Comparisons	10
<b><u>2. One-way Independent Groups ANOVA</u></b>	<b>11</b>
2.1 Basic Analysis	11
2.2 Descriptive Statistics & Graph	12
2.3 Comparisons of Means	14
2.3.1 <i>Post Hoc Comparisons</i>	14
<b>Box 1: How to Give the Results of the Significance Tests</b>	<b>15</b>
2.3.2 <i>Contrasts</i>	16
2.3.3 <i>Estimated Marginal Means</i>	18
2.4 Regression Coefficients	18
<b><u>3. One-way Independent Groups Analysis of Covariance</u></b>	<b>21</b>
3.1 Basic Analysis	21
<b>Box 2: Partial Eta Squared</b>	<b>22</b>
3.2 Testing the Homogeneity of the Regression Slopes	24
3.3 What if the Regression Slopes are Different Over the Groups?	25

<b>4. <u>Factorial Independent Groups ANOVA</u></b>	<b>27</b>
4.1 Orthogonality	27
4.2 Full Factorial Analysis	28
4.3 Testing a Main Effects Model	30
4.4 Pairwise Comparisons of Groups	31
4.4.1 Post Hoc Tests	31
4.4.2 Comparisons of Estimated Marginal Means	33
4.4.3 Holding <i>gender</i> constant	34
4.4.4 Post Hoc Tests or Estimated Marginal Means?	35
4.5 One Way of Following Up Interactions: Tests of Simple Effects (Syntax only)	35
4.6 Another Way of Following Up Interactions: Interaction Contrasts with the <i>lmatrix</i> subcommand (Syntax only)	38
<b>5. <u>Introduction to Repeated Measures Analyses</u></b>	<b>43</b>
<b>6. <u>One-way Repeated Measures (Within-subject) Analysis</u></b>	<b>45</b>
6.1 Two Levels in the Within-Subjects Factor	45
6.2 Three Levels in the Within-Subjects Factor	48
6.2.1 The Univariate and Multivariate Approaches to Repeated Measures	48
6.2.1.1 Circularity of the Variance-Covariance Matrix	48
6.2.1.2 Contrasts in Repeated Measures	49
<b>Box 3: Orthogonal Polynomial Contrasts</b>	<b>50</b>
6.2.1.3 Testing Circularity by testing Sphericity	51
6.2.1.4 Orthonormalisation	52
6.2.1.5 Sphericity	54
6.2.1.6 Epsilon -- The Extent of Departure from Circularity	55
6.2.1.7 The Multivariate Approach to Repeated Measures	55
6.2.2 Carrying out a One-way Repeated Measures ANOVA with Three Levels in the Within-Subject Factor	56
6.2.2.1 Follow-up Comparisons of Means	61

<b>7. <u>Two Within-Subjects Factors</u></b>	<b>67</b>
<b>Box 4: Using pattern rather than colour to differentiate lines in <i>GLM</i> graphs</b>	<b>70</b>
7.1 Tests of Simple Effects (syntax only)	74
<b>8. <u>A Mixed Design: One Within- and One Between-Subject Factor</u></b>	<b>77</b>
8.1 Interaction Contrasts in a Mixed Analysis using <i>contrast</i>	79
8.2 Interaction Contrasts in a Mixed Analysis using <i>lmatrix</i> and <i>mmatrix</i> (syntax only)	80
8.3 Test of Simple Effects in a Mixed Analysis (syntax only)	82
<b>9. <u>A Numeric Independent Variable in a Repeated Measures Design</u></b>	<b>83</b>
9.1 Interpreting the Effect of a Numeric Independent Variable in a Repeated Measures Analysis	83
9.2 Analysis of Covariance in Repeated Measures Designs – Adjusting the Effect of a Between-Subjects Factor	86
9.3 Analysis of Covariance in Repeated Measures Designs – Adjusting the Effect of a Within-Subjects Factor	88
9.3.1 A Within-Subject Factor and a Single Covariate	88
9.3.2 Using <i>emmeans</i> with the Covariate set to Specific Values	93
9.3.3 A Within-Subject Factor and a Varying Covariate	95
<b>10. <u>Analysing a Repeated Measures Design with Stacked Data</u></b>	<b>99</b>
<b>Box 5: The Syntax Used to Produce the Stacked Dataset</b>	<b>99</b>
10.1 <i>Mixed</i> Linear Models	102
<b>11. <u>Carrying Out Regression Analyses with GLM</u></b>	<b>103</b>
11.1 Coding Categorical Variables	103
11.2 The Equivalence of Fixed Factor ( <i>by</i> ) and Covariate ( <i>with</i> ) for a Single Dummy Variable	104
<b>12. <u>A Multiple Regression Analysis with GLM</u></b>	<b>106</b>
<b>Box 6: Specifying <i>group</i> as a <i>Fixed Factor</i> to get pairwise tests and adjusted means</b>	<b>109</b>

### **13. Interactions Among Categorical Variables** **112**

<b>Box 7: Three Important Principles</b>	<b>115</b>
--	------------

### **14. The Interaction Between a Categorical Variable and a Numeric Variable** **117**

<b>Box 8: Why Stop at the Mean?</b>	<b>120</b>
-------------------------------------	------------

14.1 Producing a Graph of the Interaction	120
---	-----

### **15. The Interaction Between Two Numeric Variables** **126**

<b>Box 9: 'Pick a point' and Regions of Significance</b>	<b>131</b>
--	------------

<b>References</b>	<b>133</b>
-------------------	------------

<b>Appendix 1: Variables in <i>glmdemo.sav</i></b>	<b>135</b>
--	------------

<b>Appendix 2: Calculation of <math>\eta^2</math> with Three IVs</b>	<b>139</b>
--	------------

<b>Appendix 3: The Matrix Multiplication in Section 6.2.1.5</b>	<b>143</b>
---	------------

## Introduction

The *GLM* (General Linear Model) in SPSS has replaced *manova*<sup>1</sup> as the point-and-click procedure for carrying out one-way, factorial, and repeated measures ANOVA and MANOVA analyses. It can also be used for regression analysis. The *manova* procedure is still available, but only with syntax (typed commands). The *manova* procedure is still very useful, because some contrasts are easier to obtain with *manova* than with *GLM*, *manova* produces discriminant function coefficients in multivariate analyses, which *GLM* doesn't (point-and-click users can get discriminant function coefficients for one-way analyses without covariates by using the *discriminant analysis* procedure), and it is possible to do canonical correlation analyses with *manova*.

By the same token, *GLM* has a few nice features of its own (apart from being accessible through point-and-click). It's easy to produce graphs of interaction effects for factorial designs, obtain means for different levels of factors adjusted for other terms in the model ('estimated marginal means') and obtain tests of pairwise simple effects.

*GLM* isn't a generalised linear model procedure – all the models fitted by *GLM* assume normally-distributed errors, unlike the logistic regression model, for example, which is a generalised linear model.

This handbook describes a few standard analyses carried out with *GLM* using both syntax and point-and-click methods. Along the way, the various features and quirks of the procedure are pointed out.

We'll use a dataset called *glmdemo.sav* for the analyses illustrated in the handout. The variables in *glmdemo.sav* are shown in the Appendix. The dataset can be downloaded from <http://www.psy.mq.edu.au/psystat/download.htm> and is also available from me at [alan.taylor@mq.edu.au](mailto:alan.taylor@mq.edu.au).

## Acknowledgements

Grateful thanks to Drs David Cairns, Anne Grunseit, Peter Langford, Lesley Inglis and Colin Wastell, and to Sue Taylor, for their encouragement, and their helpful comments on various parts of this document.

Latest update January 20th 2011

---

<sup>1</sup> In this handbook the term *manova* refers to the *manova procedure* in SPSS (which can be used for a number of different univariate and multivariate analyses), not just to multivariate analysis of variance.

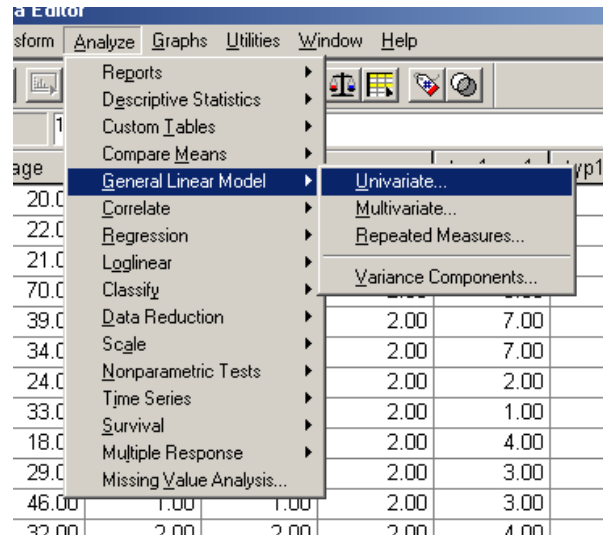




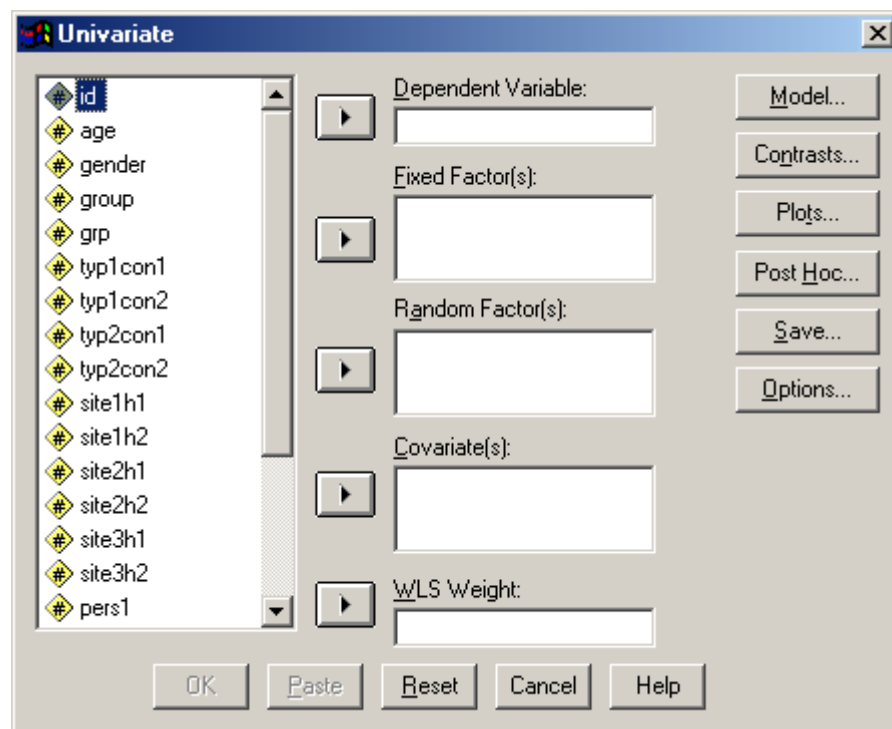
## 1. General

### 1.1 Factors and Covariates

If you click on Analyze → General Linear Model → Univariate:



you will see this panel:



The Dependent Variable slot speaks for itself. The two other slots we'll consider now are Fixed Factor(s) and Covariate(s).

### 1.1.1 Fixed Factors

Any variable entered into the *Fixed Factor(s)* slot is taken by *GLM* to be a categorical variable, for which *GLM* creates codes. (For the purposes of the regression parameters, which we'll talk about later, it uses dummy [0,1] coding. For the purposes of the ANOVA table, it uses deviation or effect [-1,0,1] coding.)

In other words, if you enter a four-category variable like *group* as a fixed factor, *GLM* will create  $4 - 1 = 3$  dummy and effect code variables for it, which are the variables actually entered into the analysis. Beware, though, if you enter a variable like *age*, which in this dataset has 27 different values: *GLM* will dutifully create 26 coding variables without so much as a message asking if you're sure you know what you're doing. The user has to be careful that variables are being treated appropriately.

### 1.1.2 Covariates

Any variable entered into the *Covariate(s)* slot is included in the analysis exactly as it is, without any sort of coding. This is the place for any equal-interval numeric variable. Forget any connotation the term 'covariate' might have as a variable which you aren't really interested in but just want to adjust or control for; a variable entered as a covariate in *GLM* may well be the independent variable of most interest to you. Bear in mind also that sometimes it may be convenient to include categorical variables with only two categories among the covariates rather than defining them as fixed factors. This way, *GLM* puts them in the model without imposing its own codes, and interpretation can be simpler, especially when you want to include dummy variables and interpret interactions (as will be discussed later, in the section on regression).

Syntax:

In *GLM* syntax, fixed factors come after *by* and covariates come after *with*. For example

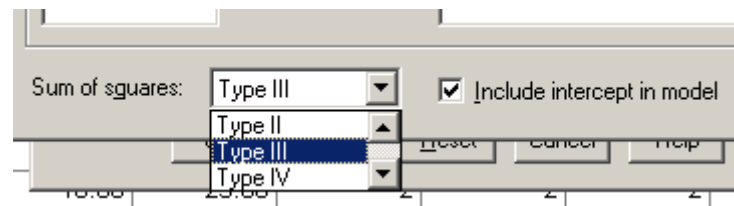
```
glm pers1 by group with age.
```

## 1.2 Sums of Squares

Users of *manova* are used to specifying either the *unique* or *sequential* method of obtaining sums of squares. Selecting *unique* means that every term in the model is tested after all other terms, as though it were entered last. This method is appropriate when the term being tested is an interaction  $a*b$  – it should be tested after  $a$  and  $b$ . We might have reservations about testing  $a$  and  $b$  with  $a*b$  held constant in an unbalanced design on the grounds that  $a*b$  should be adjusted for  $a$  and  $b$ , but not the other way around. Selecting *sequential* means that terms are tested in the order in which they are entered into the model (which is the order in which they are originally specified or the order in which they are entered on the *design* subcommand). With this method, and with  $a$  entered before  $b$ ,  $a$  is tested alone,  $b$  is tested with  $a$  held constant and  $a*b$  is tested with both  $a$  and  $b$  held constant.

*GLM* has four methods for producing sums of squares. The equivalents to *manova*'s *unique* and *sequential* are *sstype* = 3 (the default) and *sstype* = 1 respectively.

With point-and-click, *sstype* is specified by clicking on Model in the above display and then selecting the appropriate option from the pull-open menu. For most of the analyses



described here, we'll use the default *sstype* 3 (unique). Later, we'll refer to Type 2 SS, which are useful for testing main effects without adjusting for interactions.

Syntax:

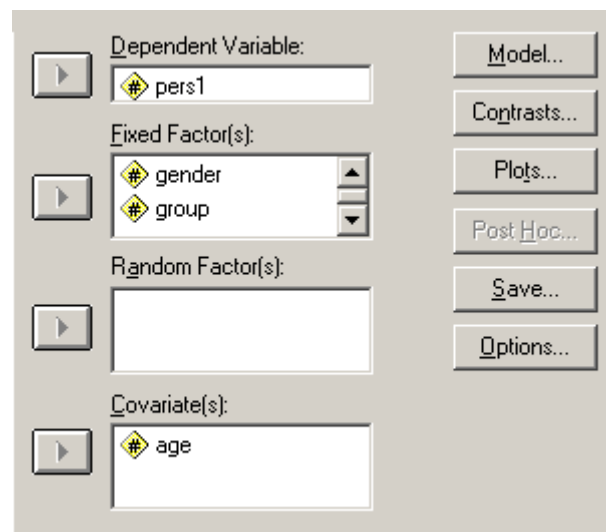
```
glm pers1 by group with age/  
method=sstype(3).
```

### 1.3 Interactions

If the variables *a* and *b* are entered as fixed factors, *GLM* automatically creates an interaction term  $a*b$ . (In these notes, the asterisk  $*$  is used as the multiplication symbol, as it is in SPSS.) For example, if *a*, *b* and *c* are entered as fixed factors, *GLM* creates  $a*b$ ,  $a*c$ ,  $b*c$  and  $a*b*c$ .

If a variable is entered as a covariate, *GLM* does not involve it in interactions (with other covariates or fixed factors) unless it is told to do so.

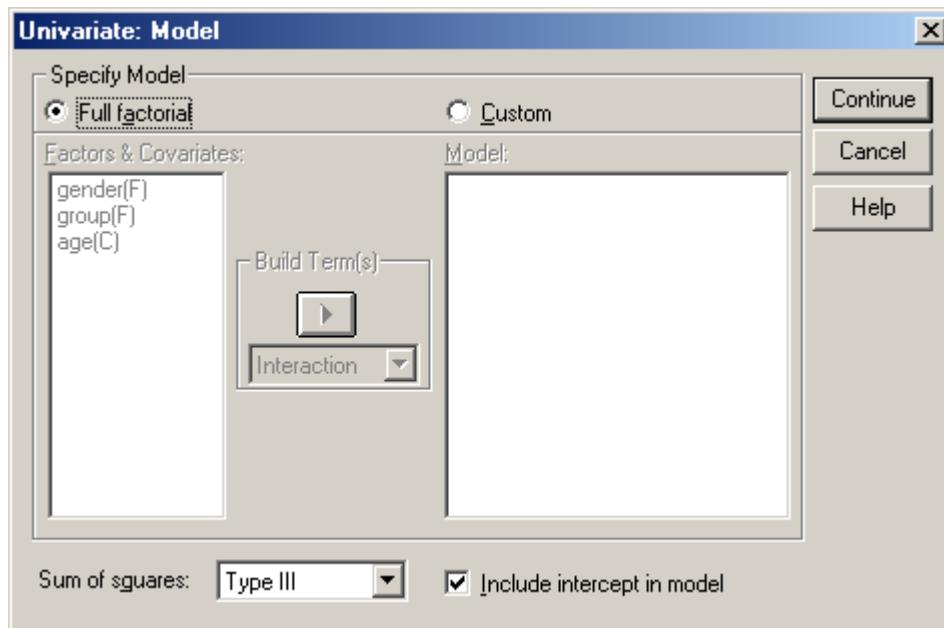
So, how do you tell *GLM* not to include interactions between fixed factors and to include interactions involving covariates? To illustrate how to do this for point-and-click, we'll start with a model containing the fixed factors *gender* and *group* and the covariate *age*:



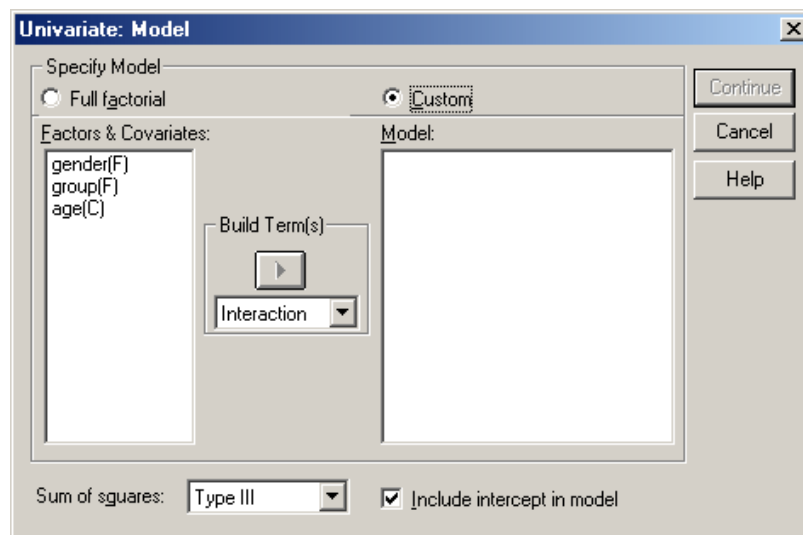
Running this analysis as is will test the model

$$pers1 = age + gender + group + gender*group.$$

In order to change the model to be tested, click on the Model button. The following display will appear:



As can be seen, the default 'Full factorial' is in force. In order to change this setting, click on the Custom radio button:



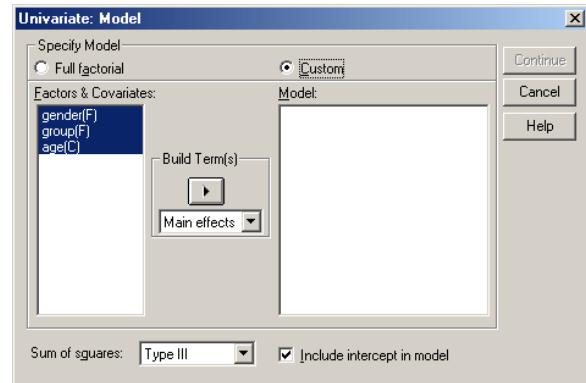
The list of variables becomes available and the user is able to 'build terms' by moving the variables from the *Factors & Covariates* list to the *Model* window. Note that the variables specified as fixed factors are followed by (F) and those specified as covariates are followed by (C).

In this example we would like to specify the following model:

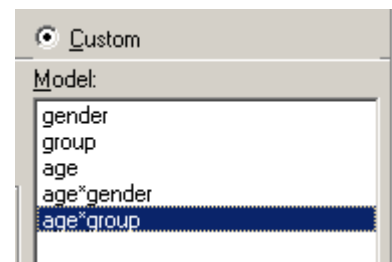
$$pers1 = age + gender + group + age*gender + age*group$$

In other words, we'd like to remove the default interaction between *gender* and *group* and include two-way interactions between *age* and both *gender* and *group*. (If we think of this design as an analysis of covariance, we're testing the assumption that the relationship between *pers1* and *age* is the same for both genders and the same across all groups.)

First, highlight *gender*, *group* and *age* in the left panel, then select *Main effects* in the pull-down *Build Term(s)* menu:



Second, click on the *Build Term(s)* arrow, and all the variables will appear in the right-hand *Model* panel. Third, select *gender* and *age* in the left-hand panel, select *Interaction* in the *Build Term(s)* menu and click on the arrow. Repeat this with *group* and *age*. The right-hand panel will look like this:



Click on *Continue* and *OK* to carry out the analysis.

#### Syntax

The default (full factorial) analysis would be carried out with

```
glm pers1 by gender group with age.
```

In syntax, custom designs are specified in the *design* sub-command:

```
glm pers1 by gender group with age/  
design=age gender group age*gender age*group.
```

### 1.4 Layout of Data

The information in this section applies to all SPSS procedures, e.g., *means*, *manova* and *t-test*, as well as to *GLM*.

#### Between-Subjects Comparisons

SPSS expects data which are going to be used to compare groups of different subjects (e.g., males versus females, older students versus younger students, subjects in the experimental group versus subjects in the control group) to be in stacked format. That is, to enable between-subject comparisons to be made, enter or arrange the data so that each subject is represented by a different line in the SPSS Data Window. For example, if you intend to compare the reaction times of 10 males and 10 females on a perceptual task, you might have a variable or column called *rt* which contains all 20 reaction times *stacked* on

top of each other (rather than being side-by-side in two columns called, for example, *rt\_f* and *rt\_m*). To tell SPSS which values of *rt* are for females and which are for males, there needs to be another variable -- *gender*, for example -- which has a code for females (e.g., 1) and a different code for males (e.g., 2). When a between-subjects analysis is carried out with *GLM*, you would tell SPSS that *rt* is the dependent variable and *gender* is the fixed factor. In the *means* procedure, *rt* would go in the *Dependent List* and *gender* in the *Independent List*. In the *Independent Samples T-Test* procedure, *rt* would go in the *Test Variable(s)* list and *gender* would be entered as the *Grouping Variable*.

### Within-Subject Comparisons

Generally, SPSS expects data which are to be used to compare subjects with themselves (e.g., repeated measures analyses, in which each subject is tested under two or more conditions) to be laid out in multivariate or side-by-side format. For example, if you wanted to compare subjects' reaction times measured before they drank a standard quantity of alcohol with their reaction times after they had drunk the alcohol, you would set up the data so that there were two variables called, for example, *pre\_rt* and *post\_rt* (or, more simply -- and simplicity and brevity in variable names are great virtues -- *rt1* and *rt2*). If there were 10 subjects, there would be 10 lines of data in the Data Window. Of course, there is no need to have a variable to identify the pre- and post-data, because the they are identified by the different variable names.

As is mentioned in Section 10 of this handbook, it is possible to carry out repeated measures analyses with stacked data in *GLM*, but this would normally be done only in exceptional circumstances. (A recent addition to SPSS, the *mixed* procedure, requires that all data be stacked, but *mixed* is beyond the scope of these notes.)

### Analyses Involving Both Between- and Within-Subject Comparisons

Many analyses involve both between- and within-subject factors. For example, we may hypothesise that the effect of alcohol (the difference between pre and post reaction times) differs for males and females. To test this hypothesis, we would need two reaction time variables (*rt1* and *rt2*, for example) and the *gender* variable. The *rt1* and *rt2* variables can be seen as the levels of a factor which might be called *time*. The test of the *time* by *gender* is a test of the hypothesis of interest.

## 1.5 A Priori and Post-hoc Comparisons

Because the aim is to illustrate the various subcommands of the *GLM* procedure, the follow-up tests carried out in the examples in this book aren't done very strictly or even very sensibly. However, the principle that *a priori* tests are treated differently from *post hoc* tests is generally adhered to. A priori tests (or contrasts or comparisons) are ones which are provoked, not by the obtained data, but purely by research questions or hypotheses. Because they are formulated ahead of time, or without reference to the obtained results, allowance has to be made only for the tests actually performed, not for all the tests that could have been done. Post hoc tests, on the other hand, while usually also bearing on the research questions and hypotheses, may be carried out in response to the obtained data. For example, with a design containing one control group and drug groups 1, 2 and 3, the researcher may not have thought of comparing drug groups 2 and 3 until she saw that the obtained means for the two groups seemed rather different. Under

these circumstances, the researcher could be seizing on what turns out to be a chance result; to try to minimise the role of chance, and keep Type I error at the nominated rate, it is important to adjust for all possible comparisons which could have been performed, not just the one that was carried out.

The advantage of thinking ahead of time and specifying a priori contrasts is that the comparisons of interest can be tested at a higher  $p$ -value (i.e., more liberal) than they could if they were part of a larger set of post hoc comparisons, especially when the number of possible contrasts is large.

### 1.6 Bonferroni Adjustment

It is sometimes argued that Bonferroni adjustment should be used for a priori tests, and some other adjustment, e.g., the Scheffé decision rule, with post hoc tests. In this handbook, Bonferroni adjustment is used freely with both a priori and post hoc tests; however, it is used for post hoc tests on the tacit understanding that only pairwise comparisons are of interest, so that the adjustment is for the number of possible pairwise post hoc comparisons; for example,  $4(4 - 1)/2 = 6$  in the example in the previous subsection. If the Type I error is to be controlled for any possible post hoc comparison, including complex ones such as the mean of the means of groups 1 and 2 versus the mean of the means of groups 3 and 4, the Scheffé method, which allows for all possible comparisons, should be used. Note, in this connection, that a significant overall test does not guarantee that any pairwise comparison will be significant; just that some contrast (possibly of no interest to the researcher) will be.

### 1.7 'Protected' Post Hoc Tests

It is conventional practice to carry out post hoc tests only after a significant overall  $F$ -test. This is sometimes called a 'protected' procedure, with the protection being against committing more Type I errors than would be expected for the nominated  $\alpha$ . In fact, as pointed out by Ramsey (2002) and others, many of the well-known post hoc procedures (the Tukey, Scheffé and the Bonferroni procedures, for example) control Type I error at the designated level in the absence of a significant overall result. Further, by using them only if the overall test is significant, the researcher is being conservative, and unduly reducing the power of the methods. For further information on this topic, see Ramsey (2002). It is worth mentioning that for users who feel uncomfortable carrying out post hoc tests in the absence of a significant overall test, but who want to have reasonable power while keeping Type I error at the nominated level, there is a protected test procedure which isn't so conservative. Hayter (1986) has shown that if pairwise  $t$ -tests are carried out following a significant  $F$ -test,  $\alpha$  will be kept at the nominated level if the differences are said to be significant if the  $t$ -values are greater than  $q_{\alpha, k-1, df}/\sqrt{2}$ , where  $q$  is the value of the studentised range statistic,  $k$  is the number of means being compared and  $df$  is the degrees of freedom of the error term. This may sound a bit obscure, but the critical value can easily be produced with SPSS. If there are four groups, each containing eight cases, so that  $k - 1 = 3$  and  $df = 28$ , and the Data Window is set up as

	groups	p	df
1	4.00	.05	28.00
2			

then the computation

```
compute cv=idf.srange(1-p,groups-1,df)/sqrt(2).
print format cv (f8.4).
list vars=p groups df cv.
```

produces the critical value as *cv*:

P	GROUPS	DF	CV
.05	4.00	28.00	2.4744

### 1.8 Orthogonal and Non-orthogonal Comparisons

The Bonferroni adjustment is applicable both when comparisons or contrasts are orthogonal and when they are non-orthogonal. Controlling Type I error is just as necessary when comparisons are independent as it is when comparisons are not independent; in fact, as Harris (1994, p. 96-103) points out, the need to apply a correction is greater when comparisons are independent than when they are not independent.

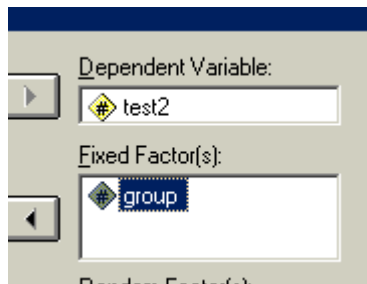
While there are some advantages in carrying out orthogonal tests in terms of avoiding redundancy, the approach adopted here is that the researcher should make the comparisons which bear on research questions; whether they are orthogonal or not is of secondary concern.



## 2. One-way Independent Groups ANOVA

### 2.1 Basic Analysis

The dependent variable in this analysis is a test score, *test2*. We want to see how *test2* varies over groups of subjects given four different dosages of a drug, as represented by *group*. With point-and-click the basic analysis is specified on the Univariate display as follows:



Syntax

```
glm test2 by group.
```

The output is as follows:

### Univariate Analysis of Variance

#### Between-Subjects Factors

		Value Label	N
GROUP	1	Control	29
	2	25 mg/kg	24
	3	50 mg/kg	31
	4	100 mg/kg	15

#### Tests of Between-Subjects Effects

Dependent Variable: TEST2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.624 <sup>a</sup>	3	1.875	3.087	.031
Intercept	2175.073	1	2175.073	3581.473	.000
GROUP	5.624	3	1.875	3.087	.031
Error	57.695	95	.607		
Total	2410.015	99			
Corrected Total	63.319	98			

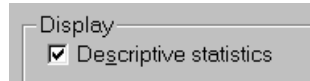
a. R Squared = .089 (Adjusted R Squared = .060)

The first panel gives the number of cases in each group. The ANOVA table shows that the test of the differences between the means is significant at  $p = .031$ . Notice that the heading of the second column tells us that Type 3 SS were used, although in this case the results for *group* wouldn't differ if another Type had been used, because there is only one

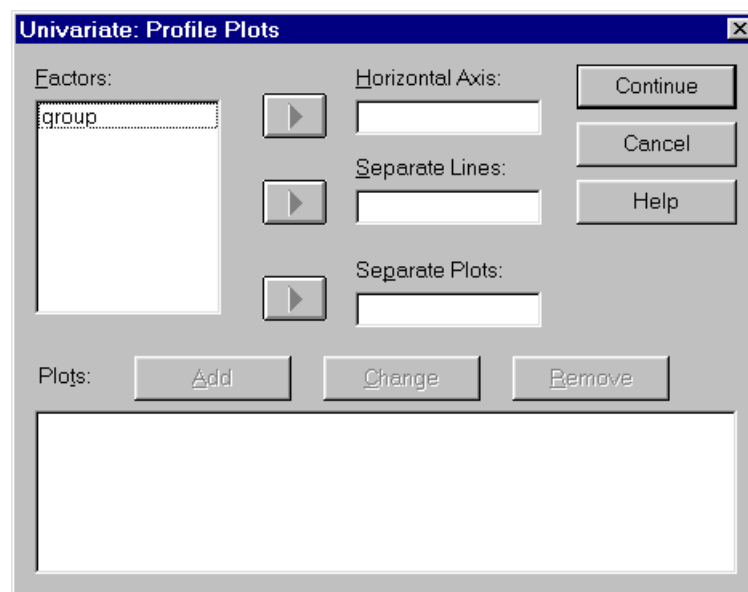
independent variable, so that the effect of *group* is not adjusted for any other variable. The  $R^2$  at the bottom of the table is equal to  $SS(\text{GROUP})/SS(\text{Total}) = 5.624/63.319$ .

## 2.2 Descriptive Statistics & Graph

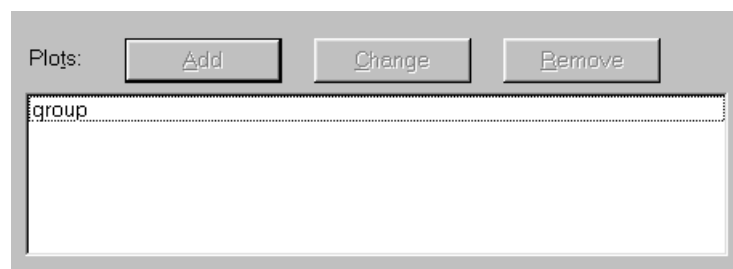
In order to obtain means and standard deviations of the dependent variable for the four groups, click on the Options button, and then check *Descriptive Statistics*:



A graph of means can be obtained by clicking on the Plots button, which will produce this display:



Graphs of means for the combinations of the levels of up to three factors can be produced. In this example, there is only one factor, *group*, so we select it as the Horizontal Axis, and then click on the Add button so that it appears in the window at the bottom of the display.



Note that when there is more than one factor, a number of different types of graphs can be produced in one analysis.

## Syntax

```
glm test2 by group/
  print=descriptives/
  plot=profile(group).
```

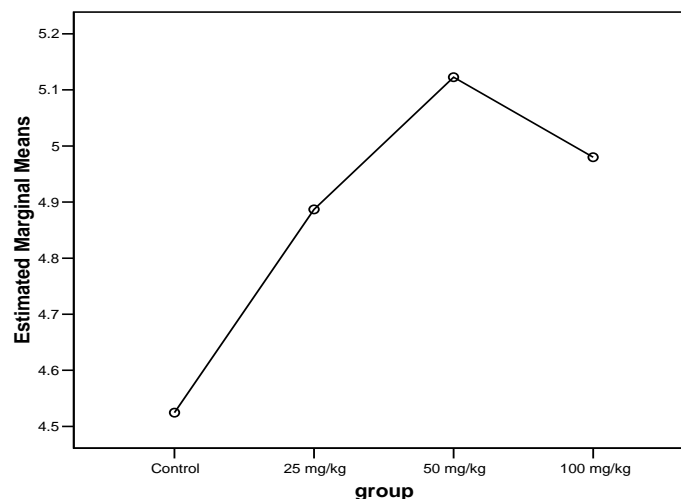
The results of these commands are shown below.

## Descriptive Statistics

Dependent Variable: TEST2

GROUP	Mean	Std. Deviation	N
1 Control	4.52	.635	29
2 25 mg/kg	4.89	.580	24
3 50 mg/kg	5.12	.814	31
4 100 mg/kg	4.98	1.159	15
Total	4.87	.804	99

## Estimated Marginal Means of Test score 2



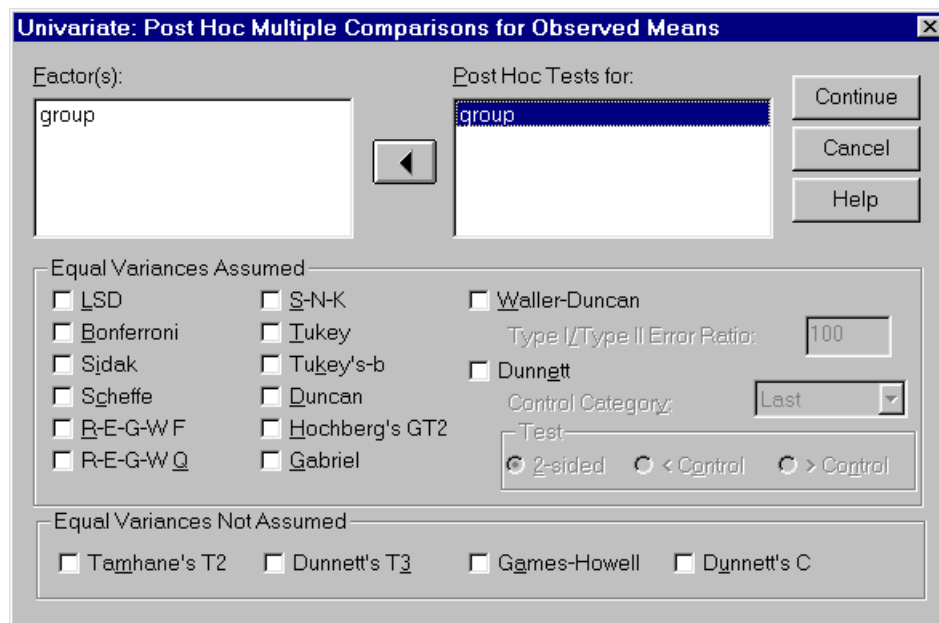
Although the means shown in the table and graph above are identical, this will not always be the case. The means in the table are not adjusted for other factors and/or covariates in the model which is being tested. The means in the graph, on the other hand, are estimated marginal means, which are actually based on the predictions of the model and, as such, are adjusted for all the variables in the model. In this example, there is only one factor, so the descriptive and estimated means are identical. If there had been another factor, or a covariate, it is unlikely that the two sets of means would have been the same. Being able to provide estimated means, both in graphs and (as will be seen later) tables, for any model, is one of the strengths of *GLM*.

## 2.3 Comparisons of Means

*GLM* offers a number of ways of carrying out comparisons of means in this simple example. (Unfortunately fewer methods are available with more complex analyses and within-subjects factors.)

### 2.3.1 *Post Hoc Comparisons*

Connoisseurs of multiple comparison tests will appreciate the plethora of possibilities offered when they click on the *Post Hoc* button and select the relevant factor, *group* in this example:



No attempt is made in this handbook to evaluate the various methods, but a few comments can be made.

- Only pairwise comparisons can be carried out with this option.
- The *LSD* option does multiple *t*-tests without making any adjustment for the number of comparisons.
- The *Bonferroni* option adjusts the *p*-values by multiplying the unadjusted *p* by the number of comparisons (the *Sidak* option is a slightly less conservative Bonferroni-type adjustment).
- In the context of pairwise tests, the *Scheffe* option is very conservative, because it allows for any contrast, not just pairwise comparisons.

Finally, an important point is that the *Post Hoc* comparisons are made only on observed means. In other words, no adjustment is made for other variables in the model. In fact, you'll notice that if a covariate is included in a point-and-click analysis, the *Post Hoc* button is no longer clickable, and if you request a post hoc test in a syntax-specified analysis which includes a covariate, the following message appears:

#### Warnings

The POSTHOC subcommand will be ignored because there are covariates in the design.

If there is more than one factor,  $a$  and  $b$ , say, the *Post Hoc* option compares the weighted means for the levels of  $a$ , say, averaged over the levels of  $b$ . This point is discussed later.

To illustrate post hoc tests, we'll check just the *LSD* and *Bonferroni* options.

#### Syntax

```
glm test2 by group/
  posthoc=group (lsd bonferroni).
```

#### Multiple Comparisons

Dependent Variable: TEST2 Test score 2

			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
(I)	(J)	Lower Bound				Upper Bound	
LSD	1 Control	2 25 mg/kg	-.36	.215	.095	-.79	.06
		3 50 mg/kg	-.60*	.201	.004	-1.00	-.20
		4 100 mg/kg	-.46	.248	.069	-.95	.04
Bonferroni	1 Control	2 25 mg/kg	-.36	.215	.572	-.94	.22
		3 50 mg/kg	-.60*	.201	.023	-1.14	-.06
		4 100 mg/kg	-.46	.248	.416	-1.12	.21

Based on observed means.

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

For the purposes of this example, only a small section of the table is shown, the part dealing with the comparison of each drug-dosage group with the control group. It can be seen that the adjusted  $p$ -values are simply the unadjusted  $p$ -values multiplied by the number of possible comparisons (six in this case); for example, the unadjusted value for the comparison between the control group and the 25 mg/kg group is .095. Multiplied by six, this is the value shown for the same comparison in the Bonferroni part of the table. If multiplying a  $p$ -value by the number of comparisons produces a value greater than one, the probability is given as 1.00.

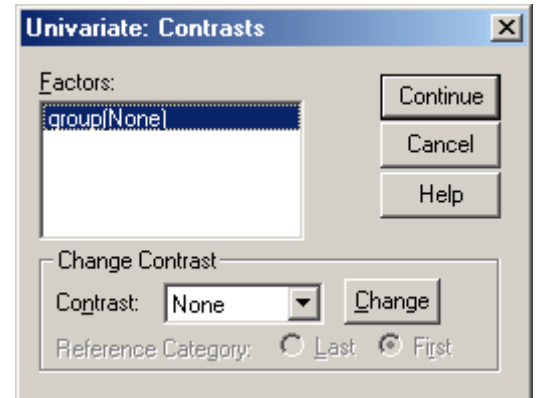
#### Box 1: How to Give the Results of the Significance Tests

The output of the *Post Hoc* results, and also the results obtained by two of the methods discussed later (contrasts and estimated marginal means), do not give a particular statistic, like  $t$  or the  $F$ -ratio. However, if you want to quote a statistic, rather than just a  $p$ -value, you can work it out for yourself from the output. The table of post hoc results above, for example, shows that the difference between the Control and 25 mg/kg group is  $-.36$ , and the standard error is  $.215$ . By double-clicking on the table, selecting the two numbers, clicking on the right mouse-button, selecting *Cell Properties*, a *Cell Format* of  $\#.\#$ , and six decimal places, we can get more accurate versions of these figures,  $.362304$  (we'll use the absolute value) and  $.215050$ . (You can use this method to get more accurate  $p$ -values, also). The difference divided by the standard error is  $1.6847$ . For the purposes of significance testing, *GLM* compares this value to a  $t$ -distribution with the degrees of freedom equal to the error term for the ANOVA, in this case 95. You could therefore quote this result as  $t(95) = 1.68, p = .095$ . If you want to quote an  $F$ -value, simply square the  $t$ -value, and give  $F(1,95) = 2.84, p = .095$ .

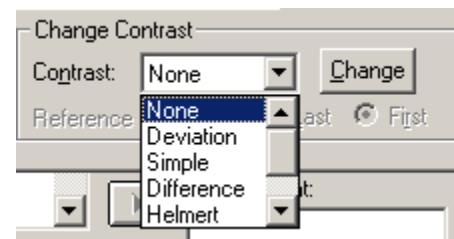
### 2.3.2 Contrasts

The *Contrast* option differs from the *Post Hoc* option in a number of ways; the first is that the contrasts are not carried out on the observed means but on the adjusted means. The second is that no adjustment can be made for the number of comparisons – *Contrast* is intended for *a priori* comparisons, where either no adjustment is made, or the users make their own adjustments. The third is that the comparisons are not limited to pairwise, but can involve more complex comparisons, such as the average of two means versus a third mean.

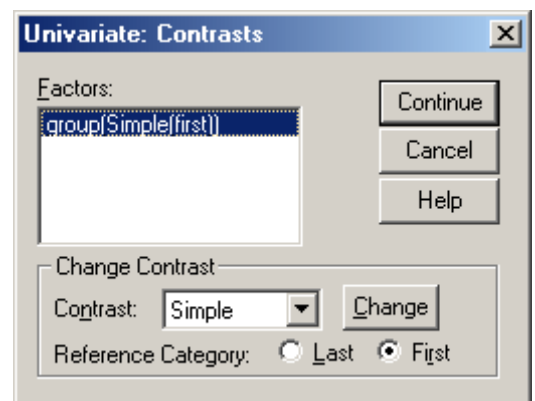
A click on the *Contrasts* button gives this display:



By default, no contrasts are tested. To select a contrast, select the factor, then pull open the *Contrast* menu:



In this example, we'll select *Simple*, in which each group is compared to one other group. Select *Simple*, then click on the *First* radio button to make the lowest-numbered group (the Control group in this case) the *Reference Category*, or the group to which each of the others is compared. Finally, click on *Change*. The display should look like this:



(For information about the various types of contrasts available, not all of which are pairwise, click on *Help* on the main SPSS menu bar (e.g., above the Data Window), then *Topics*, then *Find* and type in *contrasts*, and select the topic *Contrast Types*.)

The output for our example is shown below. In this case, because there is only one factor,

Contrast Results (K Matrix)			
			Dependent Variable
Simple Contrast <sup>a</sup>			TEST2 Test score 2
Level 2 vs. Level 1	Contrast Estimate		.362
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		.362
	Std. Error		.215
	Sig.		.095
	95% Confidence Interval for Difference	Lower Bound	-6.462E-02
		Upper Bound	.789
Level 3 vs. Level 1	Contrast Estimate		.598
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		.598
	Std. Error		.201
	Sig.		.004
	95% Confidence Interval for Difference	Lower Bound	.198
		Upper Bound	.998
Level 4 vs. Level 1	Contrast Estimate		.455
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		.455
	Std. Error		.248
	Sig.		.069
	95% Confidence Interval for Difference	Lower Bound	-3.670E-02
		Upper Bound	.947

a. Reference category = 1

the results are the same as those for the unadjusted *Post Hoc* method.

### Syntax

```
glm test2 by group/
  contrast(group)=simple(1).
```

A final point is that the *Contrast* option is the only way to test complex contrasts, such as (group 2 + group 3 + group 4)/3 – group 1, with point-and-click. For this reason it's worth exploring the *Helmert* and *Difference* contrasts<sup>2</sup> which, with a bit of juggling of the order of the groups, and a suitable choice of reference category, will enable you to do many of the complex contrasts you may want.

The syntax version of the *contrast* command is even more flexible. Say, for example, you wanted to compare (i) the overall mean of the drug-treatment groups with that for the control group, (ii) the mean of the third and fourth groups combined with the mean of the first and second groups combined and (iii) the mean of the third group with that for the first and second groups combined, ignoring the fourth group. The following syntax would do the job:

```
glm test2 by group/
  contrast(group)=special( -3 1 1 1, -1 -1 1 1, -1 -1 2 0)/
  print=test(lmatrix).
```

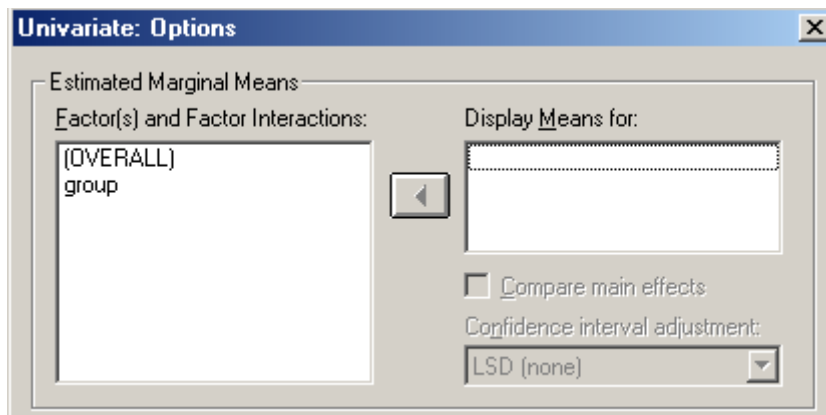
The *test(lmatrix)* option asks *GLM* to show the contrast coefficients it has used.

<sup>2</sup> In Helmert contrasts, the mean for each group is compared to the average of succeeding (higher-numbered) groups. In this case, with *group* having four levels, the three contrasts are: 3 -1 -1 -1, 0 2 -1 -1 and 0 0 1 -1. Difference contrasts are the opposite to Helmert: each group is compared to the average of the preceding groups: -1 1 0 0, -1 -1 2 0, -1 -1 -1 3. Both Helmert and Difference contrasts are orthogonal.

### 2.3.3 Estimated Marginal Means

Estimated marginal means are obtained by inserting appropriate values into a regression equation to obtain predicted values. They are especially useful for answering questions like 'how would the groups compare if all subjects had the same value on [some covariate]'. Because standard errors can be obtained for the predicted values, it is possible to assess the significance of differences between groups, while holding other variables constant.

Clicking on the *Options* button gives you access to this panel (only the top part is shown):



Using this option, all pairwise comparisons of the levels of a factor can be made, with no, or Bonferroni or Sidak, adjustments. The difference between this and the *Post Hoc* option is that the comparisons are made on the adjusted, not the observed means and, what's more, can be made among the levels of within-subject factors as well as between-subject factors. To carry out a set of Bonferroni-adjusted comparisons, select *group* and move it to the *Display Means for* window, check the *Compare main effects* box and *Bonferroni* on the *Confidence interval adjustment* pull-down menu. Because there is only one factor, the results in this case will be identical to those for the Bonferroni-adjusted *Post Hoc* option, so they are not shown.

#### Syntax

```
glm test2 by group/  
  emmeans=table(group) compare(group) adjust(bonferroni).
```

## 2.4 Regression Coefficients

You can ask for regression coefficients to be shown by clicking on the *Options* button, and selecting *Parameter Estimates*.

#### Syntax

```
glm test2 by group/  
  print=parameters.
```



The following table will be produced:

### Parameter Estimates

Dependent Variable: TEST2 Test score 2

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.980	.201	24.750	.000	4.581	5.379
[GROUP=1]	-.455	.248	-1.837	.069	-.947	3.670E-02
[GROUP=2]	-.093	.256	-.363	.718	-.602	.416
[GROUP=3]	.143	.245	.582	.562	-.344	.629
[GROUP=4]	0 <sup>a</sup>	.	.	.	.	.

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

To understand these parameters, it is important to know how *GLM* sets up the coding for categorical variables (i.e., those which are specified as fixed factors or after *by* in syntax). Dummy (0,1) variables are created for every category. For example, the dummy variable for the first group has the value of '1' for all the subjects in group 1, and zero for every other group. The dummy variable for the second group has the value of '1' for all the subjects in group 2, and zero for every other group. And so on. Because only  $k - 1$  variables are needed to represent a variable with  $k$  categories, the dummy variable for the fourth group in the above example is redundant (and would make the model impossible to deal with mathematically) so, when the model is evaluated, the dummy variable for group 4 is omitted. This means that the last (highest-numbered) category becomes the reference category. Therefore the value given in the table for intercept, 4.98, is the mean of *test2* for group 4, and each of the other coefficients shows the difference between the mean for a given group and that for group 4. For example, the difference between the mean for group 1 (4.52) and that for group 4 (4.98) is -.455.

A point to bear in mind is that because the regression coefficients for variables entered as covariates or after *with* are given without any intervention on *GLM*'s part, we could create our own dummy variables for *group*, this time making the first (lowest-numbered) group the reference category. The dummy variables could be created with the following syntax<sup>3</sup>:

```
recode group (2=1)(else=0) into g2.
recode group (3=1)(else=0) into g3.
recode group (4=1)(else=0) into g4.
```

These variables would be entered as covariates, and the *group* variable would be removed from the *Fixed Factor* slot. The table of parameters would be as shown below.

### Syntax

```
glm test2 with g2 g3 g4/
print=parameters.
```

<sup>3</sup> These commands are appropriate if there are not missing data for *group*. If there are missing values, the recode commands should be modified to avoid treating cases missing on *group* as if they are in the reference group. For example, *recode group (2=1)(missing=sysmis)(else=0) into g2*.

**Parameter Estimates**

Dependent Variable: TEST2 Test score 2

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.525	.145	31.266	.000	4.237	4.812
G2	.362	.215	1.685	.095	-6.462E-02	.789
G3	.598	.201	2.970	.004	.198	.998
G4	.455	.248	1.837	.069	-3.670E-02	.947

This time, the intercept parameter shows the mean of *test2* for group 1, and the other parameters compare the means for groups 2, 3 and 4 with that for group 1.

A final point is that because we've used *with* rather than *by*, the results in the ANOVA table are also based on our dummy coding rather than on *GLM*'s default deviation or effect coding. For this reason, the results of the significance tests are exactly the same as those shown for the parameter estimates, which wouldn't normally be the case.

**Tests of Between-Subjects Effects**

Dependent Variable: TEST2 Test score 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5.624 <sup>a</sup>	3	1.875	3.087	.031
Intercept	593.696	1	593.696	977.580	.000
G2	1.724	1	1.724	2.838	.095
G3	5.356	1	5.356	8.820	.004
G4	2.050	1	2.050	3.375	.069
Error	57.695	95	.607		
Total	2410.015	99			
Corrected Total	63.319	98			

a. R Squared = .089 (Adjusted R Squared = .060)

### 3. One-way Independent Groups Analysis of Covariance

#### 3.1 Basic Analysis

We'll now extend the previous example by taking into account a continuous numeric variable, *pers2* (a personality measure), when we compare groups. [As an exercise, you might like to run an analysis to see if the *pers2* scores differ over groups.]

With point-and-click, specify *test2* as the dependent variable, *group* as a fixed factor and *pers2* as a covariate. On *Options*, check *Descriptive statistics* and *Estimates of effect size* and de-select *Parameter estimates*. Confirm that you want *GLM* to display estimated marginal means for *group*. Click the *Plots* button, and click on Remove so that no plot is produced. For this example, we don't want to specify *Contrasts* or *Post Hoc* tests.

Syntax

```
glm test2 by group with pers2/
print=descriptives etasq/
emmeans=table(group).
```

The relevant output is as follows:

#### Descriptive Statistics

Dependent Variable: TEST2 Test score 2

GROUP	Mean	Std. Deviation	N
1 Control	4.52	.635	29
2 25 mg/kg	4.89	.580	24
3 50 mg/kg	5.12	.814	31
4 100 mg/kg	4.98	1.159	15
Total	4.87	.804	99

#### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.588 <sup>a</sup>	4	1.897	3.200	.016	.120
Intercept	35.562	1	35.562	59.981	.000	.390
PERS2	1.964	1	1.964	3.313	.072	.034
GROUP	3.467	3	1.156	1.949	.127	.059
Error	55.731	94	.593			
Total	2410.015	99				
Corrected Total	63.319	98				

a. R Squared = .120 (Adjusted R Squared = .082)

## Estimated Marginal Means

Dependent Variable: TEST2 Test score 2

	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Control	4.593 <sup>a</sup>	.148	4.300	4.887
2 25 mg/kg	4.873 <sup>a</sup>	.157	4.560	5.185
3 50 mg/kg	5.089 <sup>a</sup>	.140	4.812	5.366
4 100 mg/kg	4.941 <sup>a</sup>	.200	4.543	5.338

a. Evaluated at covariates appeared in the model: PERS2  
Personality score 2 = 4.96.

The *group* effect is no longer significant. Despite the fact that the (adjusted) means in the *Estimated Marginal Means* table are only slightly different from those given in the *Descriptive Statistics* table, *group* and *pers2* share enough variance in common with *test2* for the unique contribution of *group* to be reduced. As the original ANOVA table in Section 2.1 showed, by itself *group* accounted for  $(5.624/63.316) * 100 = 8.88\%$  of the variance of *test2*. As can be seen from the *Partial Eta Squared* column in the present ANOVA table, when *group* is adjusted for *pers2* it accounts for only  $.059 * 100 = 5.9\%$  of the total variance of *test2*. Some more information about partial  $\eta^2$  and its interpretation with different types of sums of squares is given in the following box.

### Box 2: Partial Eta Squared

The formula for partial  $\eta^2$  given in the SPSS manual is

$$\text{partial } \eta^2 = (df_h * F) / (df_h * F + df_e)$$

where  $df_h$  and  $df_e$  are the hypothesis (e.g., *group*) and error degrees of freedom respectively, and  $F$  is the  $F$ -ratio testing the hypothesis (e.g., that the effect of *group* is zero).

However, the calculation of partial  $\eta^2$  is much more easily understood by referring to the sums of squares. Given a dependent variable  $y$ , and independent variables  $a$ ,  $b$  and  $c$ ,

$$\text{partial } \eta^2 = SS_a / (SS_{\text{Total}} - [SS_b + SS_c])$$

The denominator of this expression is equal to the total sum of squares for  $y$ , minus the SS accounted for by  $b$  and  $c$ . Thus the partial  $\eta^2$  for  $y$  shows the proportion of the variance of  $y$  not explained by  $b$  and  $c$  which is explained by  $a$ .

An important point, though, is that the calculation of partial  $\eta^2$  differs according to whether Type 1 (sequential) or Type 3 (unique) SS are used. With Type 1 SS, the effect of a variable is assessed given its position in the ordered variables; for example, in a set of three variables, with Type 1 SS the effect of a variable will be different depending whether it is fitted first, second or third. With Type 3 SS, on the other hand, the effect of a variable is always assessed as if it was fitted last in a Type 1 (sequential) analysis, regardless of the order in which the variables are entered.

(contd)

In the previous analysis, the default Type 3 SS method was used. Let's repeat the analysis with Type 1 SS:

```
glm test2 by group with pers2/  
method=sstype(1)/  
print=etasq.
```

(i.e., using sequential sums of squares). The relevant part of the ANOVA table for our purposes is:

**Tests of Between-Subjects Effects**

Dependent Variable: TEST2 Test score 2

Source	Type I Sum of Squares	Partial Eta Squared
Corrected Model	7.588 <sup>a</sup>	.120
Intercept	2346.696	.977
PERS2	4.121	.069
GROUP	3.467	.059
Error	55.731	
Total	2410.015	
Corrected Total	63.319	

a. R Squared = .120 (Adjusted R Squared = .082)

Because this is a sequential analysis and *pers2* was entered before *group*, the SS for *pers2* is not adjusted for *group*: it is what it would be if *group* were not in the model. So, if we want to find out what proportion of the variance of *test2* not accounted for by *pers2* is accounted for by *group*, we would do this calculation:

$$\text{Partial } \eta^2 \text{ for } group = SS_{group} / (SS_{total} - SS_{pers2}) = 3.467 / (63.319 - 4.121) = .059.$$

It is this quantity which is given in the *Partial Eta Squared* column for *group* in our reduced table and also in the full ANOVA table.

Two important points about Type 1 SS are that variables are adjusted only for variables which precede them in the model being analysed and, because of this, Type 1 SS add up to the total SS, which is not usually the case with Type 3 SS.

What about the partial  $\eta^2$  for *pers2*? Because this is a sequential analysis, we are assessing the effect of *pers2* when it is entered first, and the effect of *group* when it is entered last. Thus the calculation is

$$\text{Partial } \eta^2 \text{ for } pers2 = SS_{pers2} / (SS_{total} - SS_{group}) = 4.121 / (63.319 - 3.467) = .069.$$

This result again agrees with what is shown in the present table, but what about the effect of *pers2* shown in the original ANOVA table, which is .034? That analysis used SS Type 3, which means that the contribution of *pers2* was assessed as if it were fitted last in a sequential analysis. In order to get the SS necessary to do the calculation, we need to do another sequential analysis, this time with *group* entered first:

(contd)

```
glm test2 by group with pers2/
method=sstype(1)/
design = group pers2.
```

[Note that to enter *group* before *pers2*, we had to use the *design* subcommand.]

The sums of squares from this analysis are:

#### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type I Sum of Squares
Corrected Model	7.588 <sup>a</sup>
Intercept	2346.696
GROUP	5.624
PERS2	1.964
Error	55.731
Total	2410.015
Corrected Total	63.319

a. R Squared = .120 (Adjusted R Squared = .082)

Using the same reasoning as we did when calculating partial  $\eta^2$  for *group*, we would expect the value for *pers2* to be  $1.964 / (63.319 - 5.624) = .034$ , and this is indeed what is shown for *pers2* in the original (Type 3 SS) ANOVA table.

Examples of the calculation of  $\eta^2$  when there are three variables in the analysis are given in Appendix 2.

### 3.2 Testing the Homogeneity of the Regression Slopes

One of the assumptions of ANCOVA is that the relationship between the covariate (*pers2* in this case) and the dependent variable (*test2*) is the same for each of the groups. That is to say, there is no interaction between the covariate and the grouping variable. As shown in Section 1.3, it is easy to test interactions in *GLM*. Refer back to Section 1.3 and, using either point-and-click or syntax, set up the commands to test the model

$$\text{test2} = \text{pers2} + \text{group} + \text{pers2} * \text{group}$$

Hint: In point-and-click, you will have to define a 'Custom' model and, with syntax, you will have to use the *design* subcommand.

The resulting ANCOVA table should look like this:

## Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

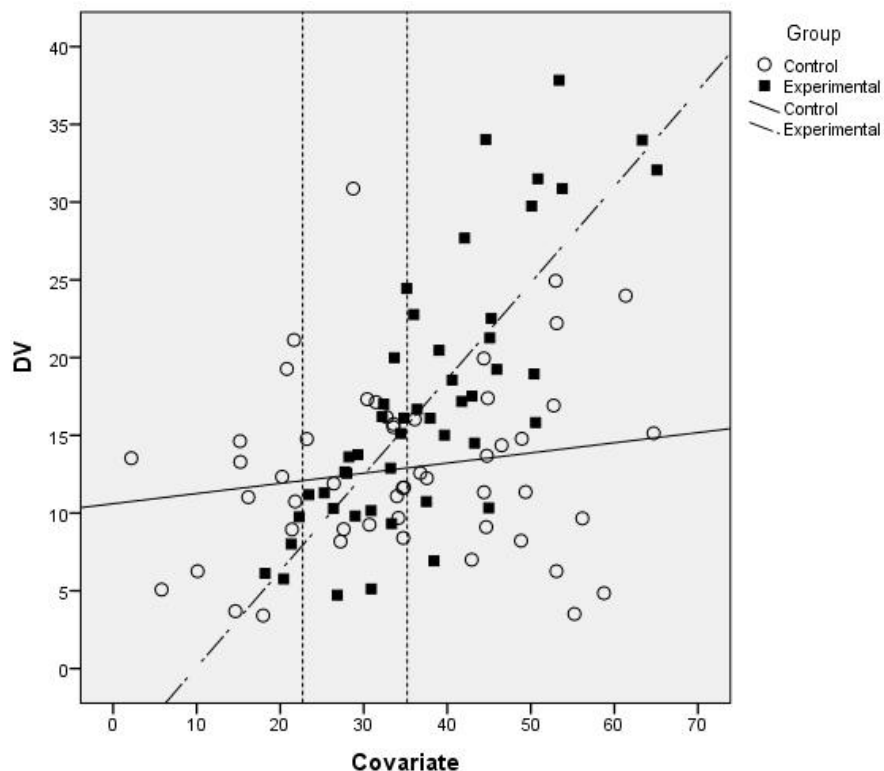
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10.226 <sup>a</sup>	7	1.461	2.504	.021
Intercept	28.866	1	28.866	49.476	.000
PERS2	1.238	1	1.238	2.122	.149
GROUP	2.358	3	.786	1.347	.264
GROUP * PERS2	2.637	3	.879	1.507	.218
Error	53.093	91	.583		
Total	2410.015	99			
Corrected Total	63.319	98			

a. R Squared = .161 (Adjusted R Squared = .097)

The interaction between *group* and *pers2* is not significant, so the assumption that the regression slopes are similar over groups is justified, and we can accept the results of the original ANCOVA analysis.

### 3.3 What if the Regression Slopes are Different Over the Groups?

If the relationship between the covariate and the dependent variable is different for two groups (say), then conventional adjustment for the covariate is inappropriate; that is, it may not be possible to make a general statement about the difference between the groups, adjusted for the covariate. All is not lost, however; it may be possible to make conditional statements such as "there is a significant difference between the groups for some values of the covariate, but not for others". Consider this graph:



Clearly there is a strong positive relationship between the covariate and the dependent variable for subjects in the experiment group, but the relationship is much less positive for subjects in the control group: the interaction effect is significant,  $F(1,95) = 37.5$ ,  $p < .0005$ . We therefore cannot make a straightforward adjustment for the covariate when comparing the groups.

Looking at the graph, however, we may feel that there could be differences between the control and experimental groups at some values of the covariate. The Johnson-Neyman technique (Johnson & Neyman, 1936; Johnson & Fay, 1950) allows us to specify the 'regions of significance' of the covariate in which there are significant differences between the groups. Using an SPSS macro provided by Hayes & Matthes (2009) we find that there is a significant difference between the groups in terms of the dependent variable (DV in the graph) when the covariate is greater than approximately 35 and less than approximately 22.5. These points are represented by the dotted vertical lines in the graph. Looking at the graph, this seems a reasonable result at the higher end, where there are a good many observations from both groups. At the lower end, however, the regression lines are clearly far apart, but there aren't many observations, especially from the experimental group. In cases like this, where the tests are based on a model fitted to the data, rather than directly on the data themselves, we have to make sure that we aren't being misled by extrapolations that go beyond the data we have actually observed.



## 4. Factorial Independent Groups ANOVA

For this analysis we'll add *gender* to *group* as a second fixed factor and use the same dependent variable, *test2*, as was used in Section 3.

### 4.1 Orthogonality

Before starting the GLM analysis, we'd like to check the 'orthogonality' of the two factors, i.e., their independence. We can do this test by creating a contingency table using the *crosstabs* procedure:

Point-and-click: *Analyze* → *Descriptive Statistics* → *Crosstabs*. Select *gender* as the Row variable and *group* as the Column variable. Click on the *Statistics* button and check *Chi-square*.

Syntax: *crosstabs gender by group/statistics=chisq*.

The relevant output is as follows:

GENDER * GROUP Crosstabulation						
Count		GROUP				
		1 Control	2 25 mg/kg	3 50 mg/kg	4 100 mg/kg	Total
GENDER	1 male	18	16	14	9	57
	2 female	11	8	17	6	42
Total		29	24	31	15	99

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.044	3	.385
Likelihood Ratio	3.038	3	.386
Linear-by-Linear Association	.723	1	.395
N of Valid Cases	99		

The table shows that the cell frequencies aren't equal, and the fact that the chi-squared statistic isn't zero (regardless of its statistical significance) means that the distribution of the cases over cells isn't as would be expected if *gender* and *group* were entirely unrelated. In this analysis, therefore, we are dealing with two correlated factors.

Note that this information is of interest because, if the factors are correlated, testing them together will mean that the effects (sums of squares and regression coefficients) of each will be adjusted for the other. In other words, we could get a different result for a given factor depending on whether it is tested by itself or with the other, correlated, factor. If the factors are not correlated, their effects (sums of squares and regression coefficients) will be the same regardless of whether they are tested together or alone. Of course,

whether the factors are tested alone or with another factor may affect the  $p$ -values, even though the sums of squares and regression coefficients are not affected. This is because the inclusion of an extra factor will reduce the variance of the error term (not necessarily significantly), which will lead the  $F$ -ratios to be larger, given the same sums of squares for effects.

Note, further, that whether the factors are correlated has no bearing on whether we may want to, or are able to, test for an interaction between the factors. In other words, it is entirely proper (and possible) to test for an interaction when the factors are not correlated and entirely proper (and possible) to test for an interaction when the factors are correlated.

## 4.2 Full Factorial Analysis

First we would like to consider a model containing *gender*, *group* and the interaction of the two variables, i.e., the full factorial model.

Point-and-click: *Analyze* → *General Linear Model* → *Univariate*. Enter *test2* as the Dependent Variable, and *gender* and *group* as Fixed Factors. Click on *Plots* and select *group* for the Horizontal Axis and *gender* for Separate Lines. Click on *Add*.

Syntax: *glm test2 by gender group/plot=profile(group\*gender)*.

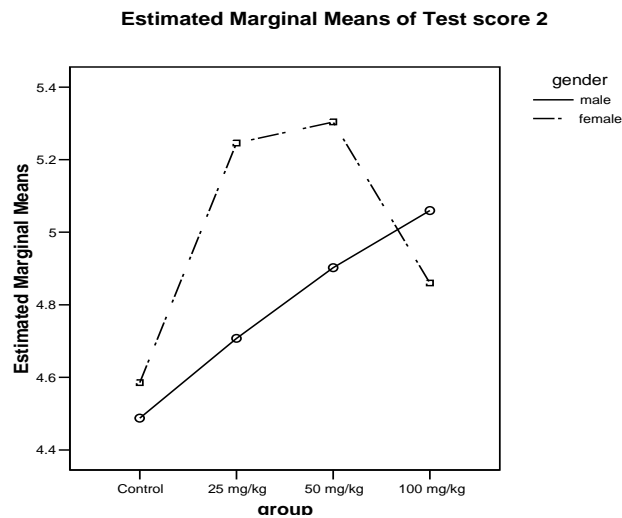
The results are:

### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.620 <sup>a</sup>	7	1.231	2.049	.057
Intercept	2065.947	1	2065.947	3437.003	.000
GENDER	.948	1	.948	1.577	.212
GROUP	5.038	3	1.679	2.794	.045
GENDER * GROUP	1.507	3	.502	.835	.478
Error	54.699	91	.601		
Total	2410.015	99			
Corrected Total	63.319	98			

a. R Squared = .136 (Adjusted R Squared = .070)

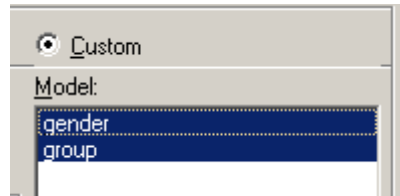


The important result here is the absence of a significant interaction between *gender* and *group* (notwithstanding the graph). If our hypothesis had been that there would be a significant interaction, our analysis might stop at this point. However, in this example we're not unduly dismayed by the lack of a significant interaction, and we'd like to continue the analysis to see what other effects have occurred.

Inspection of the ANOVA table suggests that there is a main effect for *group* but not for *gender*. One question we might have is whether we can take the results for *group* and *gender* at face value while the interaction is still in the model. Given the effect [-1, 0, 1] coding used by GLM for the ANOVA analysis, the results for *group* and *gender* would be the same whether the *gender by group* interaction was in the model or not, IF *gender* and *group* were orthogonal factors. However, as we have seen, they are not orthogonal, so the *gender* and *group* main effects are being adjusted for the interaction. There are two ways we can assess the effects of *gender* and *group* without adjusting for their interaction. One is simply to remove the interaction from the model and rerun the analysis, and this is what we'll do in the next section. Another way is to ask GLM to calculate the sums of squares on which the *F*-tests for *group* and *gender* are based in a way which does not adjust for the interaction. As we saw in Section 1.2, if Type I sums of squares are specified, GLM tests terms in the order in which they are entered into the model. If we had specified Type I SS with the commands used in the present case, *gender* would have been tested without any adjustment, *group* would have been adjusted for *gender*, and *gender by group* would have been adjusted for both *gender* and *group*. This is almost what we want, but we really want to adjust *group* for *gender* AND *gender* for *group* while not adjusting either for *gender by group*. Fortunately, GLM has another SS option which is exactly what we want: the Type II option. The Type II method does not adjust effects for other effects in which they are 'contained'. Because *gender* and *group* are both contained in the *gender by group* interaction in this sense, neither is adjusted for the interaction, but the interaction, which is not contained in *gender* or *group*, is adjusted for both. Using the Type II method is the more sophisticated way of testing the main effects in the presence of higher-order effects in a non-orthogonal design, but it's useful to know how to set up a main effects analysis to achieve much the same end, so we'll now proceed to do that.

### 4.3 Testing a Main Effects Model

To remove the interaction term from the model, go through the steps in Section 1.3 – (a) click on the *Model* button, then (b) on the *Custom* radio button, (c) select both *gender* and *group* in the *Factors & Covariates* window, (d) select *Main effects* in the *Build Terms* pull-open menu then (e) click on the *Build Terms* arrow, so that both variables appear in the right-hand *Model* window, but there is no interaction term.



The *Model* window should look like this:

#### Syntax

```
glm test2 by gender group/  
  plot=profile(group*gender)/  
  design=gender group.
```

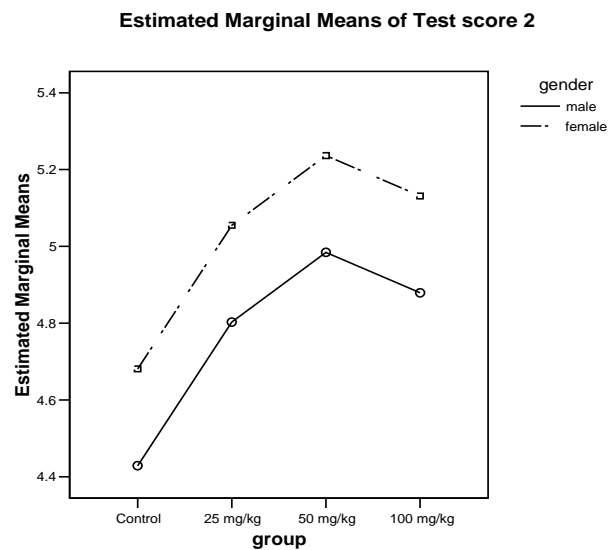
The output is:

#### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7.113 <sup>a</sup>	4	1.778	2.974	.023
Intercept	2134.395	1	2134.395	3569.619	.000
GENDER	1.489	1	1.489	2.490	.118
GROUP	4.938	3	1.646	2.753	.047
Error	56.206	94	.598		
Total	2410.015	99			
Corrected Total	63.319	98			

a. R Squared = .112 (Adjusted R Squared = .075)



Notice that the *group* effect is significant, and also that the graph of estimated marginal means faithfully reflects the fact that there is no interaction term, so that the difference between males and females is exactly the same for each group (and, equivalently, that the differences between groups are exactly the same for males and females). At this stage we could report the result for *group*, adjusted for *gender*, or remove *gender* from the model and report the unadjusted result for *group*. The appropriate action would depend on the nature of the questions being asked in the research.

#### 4.4 Pairwise Comparisons of Groups

This section illustrates the differences between *Post Hoc* tests and comparisons involving *Estimated Marginal Means*. The main difference is that the post hoc tests are conducted on means which are not adjusted for other factors or variables in the model, while estimated marginal means are adjusted (exactly what is meant by 'adjusted' will be discussed in the section on estimated marginal means). In other words, the group means compared by the post hoc tests in this analysis are exactly the same as those which would be compared if *gender* was not included in the model (although the *p*-values are slightly different, because of the effect of the inclusion of *gender* on the error term).

##### Syntax

```
glm test2 by gender group/
  print=descriptives parameters/
  emmeans=table(group) compare(group)/
  emmeans=table(gender*group)/
  posthoc=group(lsd)/
  design=gender group.
```

##### 4.4.1 Post Hoc Tests

Below is the table of means produced by the *print=descriptives* subcommand (the means

**Descriptive Statistics**

Dependent Variable: TEST2 Test score 2

GENDER	GROUP	Mean	Std. Deviation	N
1 male	1 Control	4.4874	.540	18
	2 25 mg/kg	4.7074	.535	16
	3 50 mg/kg	4.9022	.589	14
	4 100 mg/kg	5.0599	1.114	9
	Total	4.7414	.683	57
2 female	1 Control	4.5855	.792	11
	2 25 mg/kg	5.2460	.520	8
	3 50 mg/kg	5.3040	.939	17
	4 100 mg/kg	4.8601	1.320	6
	Total	5.0414	.924	42
Total	1 Control	4.5246	.635	29
	2 25 mg/kg	4.8869	.580	24
	3 50 mg/kg	5.1225	.814	31
	4 100 mg/kg	4.9800	1.159	15
	Total	4.8687	.804	99

and differences in this and the following tables are shown to four decimal places). Note that the means shown in the *Total* section of the table are weighted means; i.e., they reflect the numbers of males and females which are combined to produce the total mean. For example, the mean for the 25 mg/kg group = (the mean for males in the 25 mg/kg group \* the number of males in that group + the mean for females in the 25 mg/kg group \* the number of females in that group) / the total number of males and females in the 25 mg/kg group =  $(4.7074 * 16 + 5.2460 * 8) / (16 + 8) = 4.8869$ . If the mean for subjects in the 25 mg/kg group were calculated as (the mean for males in the 25 mg/kg group + the mean for females in the 25 mg/kg group) / 2 =  $(4.7074 + 5.2460) / 2 = 4.9767$ , this would be the unweighted mean, i.e., the male and female means are weighted equally in the calculation, even though there are twice as many males as females. Of course, if the number of males and females had been equal, the weighted and unweighted means would be the same.

The *Post Hoc* tests in *GLM* are performed on observed, unadjusted, weighted means. Below is part of the table of post hoc significance tests produced by the above commands.

**Multiple Comparisons**

Dependent Variable: TEST2 Test score 2

LSD

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.
1 Control	2 25 mg/kg	-.3623	.213	.093
	3 50 mg/kg	-.5979	.200	.004
	4 100 mg/kg	-.4553	.246	.067

Based on observed means.

The differences shown in the table can be derived from the 'total' part of the table of means above. For example, the difference between the Control and 25 mg/kg group is  $4.5246 - 4.8869 = -.3623$ , as shown in the table of multiple comparisons.

The note at the bottom of the table above – *Based on observed means* – reminds us that the group means compared by *Post Hoc* tests are not adjusted for *gender*. If we had included *gender* in this main effects analysis in order to 'take account of', or 'adjust for', the fact that the proportion of males and females differs over groups, we would not use *Post Hoc* tests.

#### 4.4.2 Comparisons of Estimated Marginal Means

Estimated marginal means are derived from the model which is being tested, and therefore take other factors and variables into account when testing each factor or variable. Part of the table of parameter estimates for the present (main effects) model is shown below. These estimates can be used to calculate means for the various cells in the design; for example, the mean for the males in the control group is equal to  $5.1312 + 1 * -.2520 + 1 * -.4501 = 4.4291$ . The mean for females in the same group is  $5.1312 + 0$

##### Parameter Estimates

Dependent Variable: TEST2 Test score 2

Parameter	B
Intercept	5.1312
[GENDER=1]	-.2520
[GENDER=2]	0 <sup>a</sup>
[GROUP=1]	-.4501
[GROUP=2]	-.0762
[GROUP=3]	.1052
[GROUP=4]	0 <sup>a</sup>

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

$* -.2520 + 1 * -.4501 = 4.6811$ . These means are not the same as the observed means in the above table -- we have 'degraded' the model by removing the interaction, so, unless the interaction is exactly zero, the 'unsaturated' model cannot reproduce the original values.

The estimates shown in the table below, produced by the *emmeans=table(gender\*group)*

Dependent Variable: TEST2 Test score 2

		Mean
1 male	1 Control	4.4290
	2 25 mg/kg	4.8029
	3 50 mg/kg	4.9843
	4 100 mg/kg	4.8792
2 female	1 Control	4.6811
	2 25 mg/kg	5.0550
	3 50 mg/kg	5.2364
	4 100 mg/kg	5.1312

subcommand, are all derived from the model parameter estimates as shown above. The *group* means shown in the next table, produced by the subcommand *emmeans=table(group) compare(group)*, are derived from the estimated marginal means for male and female subjects using the unweighted method; for example, the overall

**Estimates**

Dependent Variable: TEST2 Test score 2

	Mean
1 Control	4.5550
2 25 mg/kg	4.9289
3 50 mg/kg	5.1103
4 100 mg/kg	5.0052

mean for the 25 mg/kg group =  $(4.8029 + 5.055) / 2 = 4.929$ , which is the same as that given in the above table allowing for rounding error.

Finally, the comparisons produced by the subcommand *emmeans=table(group) compare(group)* are based on the unweighted means. Part of the table is shown below:

**Pairwise Comparisons**

Dependent Variable: TEST2 Test score 2

(I)	(J)	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>
1 Control	2 25 mg/kg	-.3739	.214	.083
	3 50 mg/kg	-.5553	.202	.007
	4 100 mg/kg	-.4501	.246	.070

Based on estimated marginal means

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

To take one example, the difference of -.3739 between the Control and 25 mg/kg group is equal to  $4.555 - 4.9289$ , values shown in the previous table.

**4.4.3 Holding *gender* constant**

Although we've used the regression equation to calculate the means for the eight *gender* \* *group* cells, then used the unweighted means of the male and female results to compare the groups, we can arrive at the same results by substituting one value for *gender* in the equation and directly calculating the mean for each group. We want males and females to be equally weighted in the calculation, so we'll 'split' the dummy code for *gender* into two equal parts, giving rise to .5. We can then calculate the mean for each group with the following:

Mean for Group 1 =  $5.1312 + .5 * -.252 + 1 * -.4501 = 4.5551$

Mean for Group 2 =  $5.1312 + .5 * -.252 + 1 * -.0762 = 4.9290$

Mean for Group 3 =  $5.1312 + .5 * -.252 + 1 * .1052 = 5.1103$

Mean for Group 4 =  $5.1312 + .5 * -.252 = 5.0052$

Within rounding error, these are the values given in the table supplied by the *Estimated marginal means* command, which are also those obtained by calculating the unweighted means of the predicted values for male and female subjects in each group.



What is meant by 'adjusting for', 'taking account of' or 'holding constant' a variable *A* when testing another variable *B*, then, is that the means which are to be compared for *B* are obtained from a regression equation which contains a constant value of *A*, along with the codes appropriate for the different levels of *B*.

It is worth noting that, if we wanted to obtain the estimated marginal means for males and females, adjusting for *group*, the same method would apply, this time using the dummy codes for females (0) and males (1), and 1/4 as the value for each of the groups in the regression equations:

Mean for females =  $5.1312 + 0 * -.2520 + 1/4 * -.4501 + 1/4 * -.0762 + 1/4 * .1052$ .

Mean for males =  $5.1312 + 1 * -.2520 + 1/4 * -.4501 + 1/4 * -.0762 + 1/4 * .1052$ .

#### 4.4.4 Post Hoc tests or Estimated Marginal Means?

The choice between the *Post Hoc* method and the *Estimated Marginal Means* method is clear-cut if you want to adjust for other factors or variables in the analysis – don't use the *Post Hoc* method, and do use either the *Estimated Marginal Means* method, or the *contrasts* subcommand, or the parameter estimates, both of which are also based on adjusted means. In cases where there are covariates (i.e., numeric variables) in the model, there is likewise no choice -- *Post Hoc* is not available when there are covariates. In other cases, such as one-way ANOVAs, the two methods will give identical results. My advice is, when in doubt, use *Estimated Marginal Means*.

### 4.5 One Way of Following Up Interactions: Tests of Simple Effects (Syntax only)

The full factorial analysis in Section 4.2 showed that the interaction between *gender* and *group* wasn't significant, but for the purposes of demonstration we'll act as if it was.

One of the things that researchers may want to do when they discover an interaction between fixed factors is to carry out tests of simple effects. In the present case, tests of the simple effects of *gender* and *group* involve comparing males and females within each group (four comparisons in all) and comparing groups separately for males and females (two comparisons if we just test the overall effect of *group* for males and females, but 12 comparisons if we do every pairwise comparison of the four groups separately for each gender)

*GLM* allows the testing of simple effects, but only with syntax (with point-and-click, you can obtain the means required for the tests of simple effects by clicking on *Options*, and specifying *gender\*group* in the *Display Means for* window, but no significance tests are done). The syntax for the present example is:

```
glm test2 by gender group/
  emmeans=table(gender*group) compare(group)/
  emmeans=table(gender*group) compare(gender).
```

The output is as follows:

## Estimated Marginal Means

### Estimates

Dependent Variable: TEST2 Test score 2

		Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1 male	1 Control	4.487	.183	4.124	4.850
	2 25 mg/kg	4.707	.194	4.322	5.092
	3 50 mg/kg	4.902	.207	4.491	5.314
	4 100 mg/kg	5.060	.258	4.547	5.573
2 female	1 Control	4.586	.234	4.121	5.050
	2 25 mg/kg	5.246	.274	4.702	5.790
	3 50 mg/kg	5.304	.188	4.930	5.677
	4 100 mg/kg	4.860	.317	4.231	5.489

### Pairwise Comparisons

Dependent Variable: TEST2 Test score 2

			Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
(I)	(J)	Lower Bound				Upper Bound	
1 male	1 Control	2 25 mg/kg	-.220	.266	.411	-.749	.309
		3 50 mg/kg	-.415	.276	.137	-.964	.134
		4 100 mg/kg	-.572	.317	.074	-1.201	5.626E-02
	2 25 mg/kg	1 Control	.220	.266	.411	-.309	.749
		3 50 mg/kg	-.195	.284	.494	-.758	.369
		4 100 mg/kg	-.352	.323	.278	-.994	.289
	3 50 mg/kg	1 Control	.415	.276	.137	-.134	.964
		2 25 mg/kg	.195	.284	.494	-.369	.758
		4 100 mg/kg	-.158	.331	.635	-.816	.500
	4 100 mg/kg	1 Control	.572	.317	.074	-5.626E-02	1.201
		2 25 mg/kg	.352	.323	.278	-.289	.994
		3 50 mg/kg	.158	.331	.635	-.500	.816
2 female	1 Control	2 25 mg/kg	-.660	.360	.070	-1.376	5.512E-02
		3 50 mg/kg	-.718*	.300	.019	-1.314	-.123
		4 100 mg/kg	-.275	.393	.487	-1.056	.507
	2 25 mg/kg	1 Control	.660	.360	.070	-5.512E-02	1.376
		3 50 mg/kg	-5.795E-02	.332	.862	-.718	.602
		4 100 mg/kg	.386	.419	.359	-.446	1.218
	3 50 mg/kg	1 Control	.718*	.300	.019	.123	1.314
		2 25 mg/kg	5.795E-02	.332	.862	-.602	.718
		4 100 mg/kg	.444	.368	.231	-.287	1.175
	4 100 mg/kg	1 Control	.275	.393	.487	-.507	1.056
		2 25 mg/kg	-.386	.419	.359	-1.218	.446
		3 50 mg/kg	-.444	.368	.231	-1.175	.287

Based on estimated marginal means

\*. The mean difference is significant at the .050 level.

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

**Estimates**

Dependent Variable: TEST2 Test score 2

		Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1 male	1 Control	4.487	.183	4.124	4.850
	2 25 mg/kg	4.707	.194	4.322	5.092
	3 50 mg/kg	4.902	.207	4.491	5.314
	4 100 mg/kg	5.060	.258	4.547	5.573
2 female	1 Control	4.586	.234	4.121	5.050
	2 25 mg/kg	5.246	.274	4.702	5.790
	3 50 mg/kg	5.304	.188	4.930	5.677
	4 100 mg/kg	4.860	.317	4.231	5.489

**Pairwise Comparisons**

Dependent Variable: TEST2 Test score 2

			Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
(I)	(J)	Lower Bound				Upper Bound	
1 Control	1 male	2 female	-9.811E-02	.297	.742	-.687	.491
	2 female	1 male	9.811E-02	.297	.742	-.491	.687
2 25 mg/kg	1 male	2 female	-.539	.336	.112	-1.205	.128
	2 female	1 male	.539	.336	.112	-.128	1.205
3 50 mg/kg	1 male	2 female	-.402	.280	.155	-.958	.154
	2 female	1 male	.402	.280	.155	-.154	.958
4 100 mg/kg	1 male	2 female	.200	.409	.626	-.612	1.011
	2 female	1 male	-.200	.409	.626	-1.011	.612

Based on estimated marginal means

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

The first pair of tables shows the results for the tests of the simple effects of *group* and the second pair of tables shows the results for *gender*. The first table in each pair shows the means and the second shows the differences between the means and the results of pairwise comparisons. There's a fair bit of redundancy in the second table of each pair: *GLM* not only compares the Control group with the 25 mg/kg group, but also the 25 mg/kg group with the Control group. We can't complain that it doesn't cover all the possibilities. The *p*-values for the tests shown here are unadjusted. If these were all *post hoc* tests, we might want to obtain a critical *p*-value for the *group* pairwise comparisons by dividing .05 by 12 and for the *gender* comparisons by dividing .05 by 4. On the other hand, if only some of the comparisons were relevant, and they were to be treated as *a priori* tests, .05 might be divided only by the number of actual comparisons (rather than all that could be done), or used as is, if *alpha* were to be maintained on a per-decision basis.

As discussed in Section 2.3.1, you can obtain a *t*-value to quote by dividing the *Mean Difference* given in the tables by the *Std. Error*, and using the relevant degrees of freedom. In the present case, for example, the result for the comparison of the Control and 25 mg/kg groups would be  $t = .220/.266$ , and the degrees of freedom would be that for the error for the *gender\*group* interaction, 91:  $t(91) = .83, p = .411$ .

#### 4.6 Another Way of Following Up Interactions: Interaction Contrasts with the *lmatrix* subcommand (Syntax only)

Interactions are all about differences between differences. If we had an interaction between gender and a group factor with only two levels, we would know that the difference between groups was different for males and females (or, equivalently, that the difference between males and females was different for the two groups). If an interaction occurs between factors with more than two levels, we may want to investigate further to find out exactly where the differences between differences are occurring. In the present example we're acting as if a significant interaction occurred between *gender* and *group*. Because there are four groups, we may want to ask such questions as 'Is the difference between the Control and 25 mg/kg groups different for males and females?' 'Is the difference between the Control and 50 mg/kg groups different for males and females?' 'And so on. Or, equivalently, we might ask whether the difference between males and females is different for the Control and 100 mg/kg groups, and so on.

The only way we can answer these questions in *GLM* is to specify them in syntax using the *lmatrix* subcommand. This is a not entirely straightforward command to use, but it helps if you understand how the *GLM* model is set up.

As mentioned in Section 2.4 on regression coefficients, the *GLM* model is set up with a variable representing every level of each fixed factor. The full factorial model we

Main effects						Interactions							
m	f	cont	25	50	100	m	m	m	m	f	f	f	f
c	25	50	100	c	25	50	100	c	25	50	100	c	25
Codes for (mc – m25) – (fc – f25):						1	-1	0	0	-1	1	0	0
Codes for (m25 – f25) – (m100 – f100):						0	1	0	-1	0	-1	0	1
Codes for (mc – (m25 + m50 + m100)/3) - (fc – (f25 + f50 + f100)/3)						1	-1/3	-1/3	-1/3	-1	1/3	1/3	1/3

have been testing can be represented as in the above table. In the first part of the table are the columns representing the main effects, *gender* first (males, numbered '1', first, then females, numbered '2'), then *group*, with the four levels of treatment in the order of their numerical codes. The interaction terms in the second part of the table can then be created by combining, first, males with each level of *group*, then females with each level of *group*. All the possible combinations are represented. (If *group* had come before *gender* in the model, the order of the interactions would be Control-Male, Control-Female, 25 mg/kg-Male, 25 mg/kg-Female, etc).

Now let's consider a possible interaction contrast: Is the difference between the Control and 25 mg/kg group different for males and females? Using the abbreviations from the table, we can express this question as follows. Is

$$(mc - m25) - (fc - f25)$$

different from zero? If the expression is simplified<sup>4</sup> to get rid of the brackets, it becomes

$$mc - m25 - fc + f25,$$

then, if we introduce some coefficients, all with the value of '1', we get

$$1 * mc - 1 * m25 - 1 * fc + 1 * f25.$$

The '1's don't change anything about the expression, but we can use them to tell *GLM* what sign the variables they are associated with are to have when the contrast is performed. The pattern of '1's is shown in the above table. Notice that the interaction columns which aren't involved in this contrast have been assigned a zero and will be ignored in the calculation.

We will now ask two more questions about the interaction before using the coefficients to specify a test in *GLM*. The first question is, 'is the difference between males and females in the 25 mg/kg group different from the difference between males and females in the 100 mg/kg group? In other words, is  $(m25 - f25) - (m100 - f100)$  different from zero? Simplifying, we get  $m25 - f25 - m100 + f100$ . The appropriate coefficients are shown in the above table.

The second question is a bit more complicated: Is the difference between the control and the average of the three treatment groups different for males and females? In other words, is

$$(mc - (m25 + m50 + m100)/3) - (fc - (f25 + f50 + f100)/3)$$

different from zero?

This expression simplifies to:

$$mc - m25/3 - m50/3 - m100/3 - fc + f25/3 + f50/3 + f100/3,$$

so the coefficients are

$$1 - 1/3 - 1/3 - 1/3 - 1 + 1/3 + 1/3 + 1/3$$

and they have been placed in the table above.

The syntax for implementing these contrasts is

```
glm test2 by gender group/
print=descriptives/
lmatrix="g*gp contrasts" gender*group 1 -1 0 0 -1 1 0 0;
gender*group 0 1 0 -1 0 -1 0 1; gender*group 1 -1/3 -1/3 -1/3 -1 +1/3 +1/3 + 1/3.
```

---

<sup>4</sup> Once the expression is simplified, it can rearranged to this form:  $(mc - fc) - (m25 - f25)$ . This expression is asking whether the difference between males and females is different for different groups, which is different from the question asked originally, but which has exactly the same answer, and is an equivalent question.

The *lmatrix* subcommand is followed by a title in inverted commas which is used in the output, then the specification of the interaction, so that *GLM* knows which effect the contrast coefficients are referring to, then the coefficients themselves, followed by a semi-colon if there is another set of coefficients to come. Although we've limited ourselves to testing one contrast per degree of freedom  $[(2 - 1) * (4 - 1) = 3]$ , *GLM*, unlike the *manova* procedure, places no limit on the number of contrasts that can be specified in a given analysis.

The relevant tables in the output are:

#### Descriptive Statistics

Dependent Variable: TEST2 Test score 2

GENDER	GROUP	Mean	Std. Deviation	N
1 male	1 Control	4.4874	.540	18
	2 25 mg/kg	4.7074	.535	16
	3 50 mg/kg	4.9022	.589	14
	4 100 mg/kg	5.0599	1.114	9
	Total	4.7414	.683	57
2 female	1 Control	4.5855	.792	11
	2 25 mg/kg	5.2460	.520	8
	3 50 mg/kg	5.3040	.939	17
	4 100 mg/kg	4.8601	1.320	6
	Total	5.0414	.924	42
Total	1 Control	4.5246	.635	29
	2 25 mg/kg	4.8869	.580	24
	3 50 mg/kg	5.1225	.814	31
	4 100 mg/kg	4.9800	1.159	15
	Total	4.8687	.804	99

**Contrast Results (K Matrix)<sup>a</sup>**

		Dependent Variable
Contrast		TEST2 Test score 2
L1	Contrast Estimate	.440
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	.440
	Std. Error	.448
	Sig.	.328
	95% Confidence Interval for Difference	
	Lower Bound	-.449
	Upper Bound	1.330
L2	Contrast Estimate	-.738
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	-.738
	Std. Error	.529
	Sig.	.166
	95% Confidence Interval for Difference	
	Lower Bound	-1.789
	Upper Bound	.312
L3	Contrast Estimate	.149
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	.149
	Std. Error	.358
	Sig.	.678
	95% Confidence Interval for Difference	
	Lower Bound	-.561
	Upper Bound	.859

a. Based on the user-specified contrast coefficients (L') matrix: g\*gp contrasts

The result of the first contrast (second table) is .440. Using the means in the first table, and the contrast coefficients we worked out, the calculation is as follows:

$$1 * 4.4874 - 1 * 4.7074 + 0 * 4.9022 + 0 * 5.0599 - 1 * 4.5855 + 1 * 5.246 + 0 * 5.304 + 0 * 4.8601 = .440.$$

The third contrast can be checked in the same way:

$$1 * 4.4874 - 1/3 * 4.7074 - 1/3 * 4.9022 - 1/3 * 5.0599 - 1 * 4.5855 + 1/3 * 5.246 + 1/3 * 5.304 + 1/3 * 4.8601 = .149.$$

*GLM* seems to have got it right! Of course, we could work out these results ourselves, but *GLM* also provides significance tests, and it's these results we're interested in. In this case, as would be expected, given the lack of a significant overall test of the interaction, none of the contrasts is significant.





## 5. Introduction to Repeated Measures Analyses

In repeated measures analyses, each subject is measured more than once under different conditions. For example, a subject may be asked to detect a signal under varying levels of illumination, or be given a series of written vignettes and asked to rate some aspect of the behaviour described in each vignette on a common scale. In the second example, the different vignettes are the different conditions, and the responses on the common scale are the repeated measures. In the classic repeated measures design, a pretest measure is obtained, subjects are treated in one or more different ways, measured again (posttest) then perhaps measured some time later (followup).

The aim of a repeated measures analysis is to compare responses made under the different conditions of the within-subjects factor and, if there is also a between-subjects factor, to see whether any differences between levels of the within-subjects factor vary over the levels of the between-group factor (experimental versus control, for example).

*GLM*'s main repeated measures procedure expects data to be in what is called *multivariate* layout. That is, each of the levels of a within-subjects factor is represented by a different variable. For instance, if there is a pretest, posttest and followup measure, there will be three variables, which might be called *pre*, *post* and *fu*. If there are two within-subjects factors, there will usually be a variable for each combination of conditions. In one of the examples we'll be looking at later, each subject has measurements taken at three sites on their skull on both hemispheres, six measurements in all. The variables are *site1h1*, *site1h2*, *site2h1*, *site2h2*, *site3h1* and *site3h2*.

One of the disadvantages of the multivariate layout is that if a subject is missing on one or more of a number of the variables making up the levels of the within-subject factor(s), he or she is omitted from the analysis (it may be reasonable to impute one or more values in order to avoid this). *GLM* also handles another kind of layout, *stacked* (also referred to as *mixed*), in which the absence of a particular measurement doesn't have such drastic consequences. In this layout, all the measurements for a given subject make up one variable, and there are other variables which show which condition a given measurement is from. In the two-factor example above, there would be a variable showing the results of measurements on the subjects, a variable for site (with three codes, 1, 2 and 3, say) and another for hemisphere (two codes). As a consequence, there are as many 'cases' in the dataset for each subject as there are observations for that subject. The omission of the observation for one combination of conditions (site 2, left hemisphere, for example) does not result in the removal of all of that subject's data from the analysis. An example of an analysis of stacked data will be described later.



## 6. One-way Repeated Measures (Within-subject) Analysis

### 6.1 Two Levels in the Within-Subjects Factor

The simplest repeated measures design has just two measurements – so there are two levels in the within-subject factor. The data from this design can be analysed using a paired *t*-test. For example:

Point-and-click: *Analyze* → *Compare Means* → *Paired Samples T-Test*. Select *test1* and *test2* and click on *OK*.

Syntax: *t-test pairs=test1 test2*.

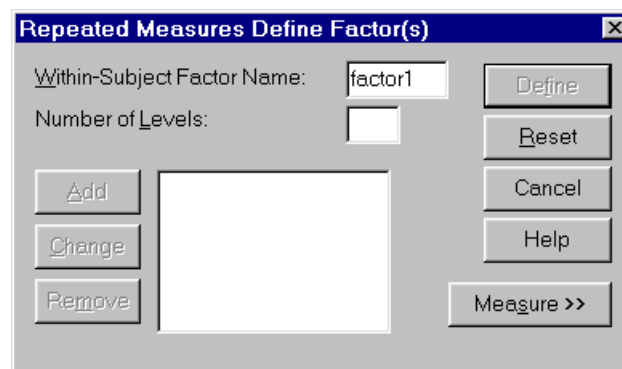
**Paired Samples Statistics**

			Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TEST1	Test score 1	4.61	99	.794	.080
	TEST2	Test score 2	4.87	99	.804	.081

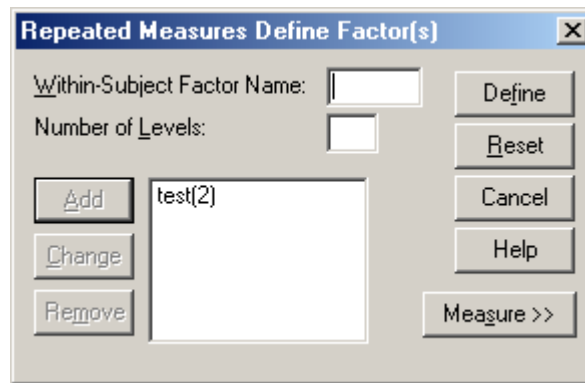
**Paired Samples Test**

			Paired Differences					
			Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Pair 1	TEST1	Test score 1 - TEST2	-.25	1.076	.108	-2.345	98	.021

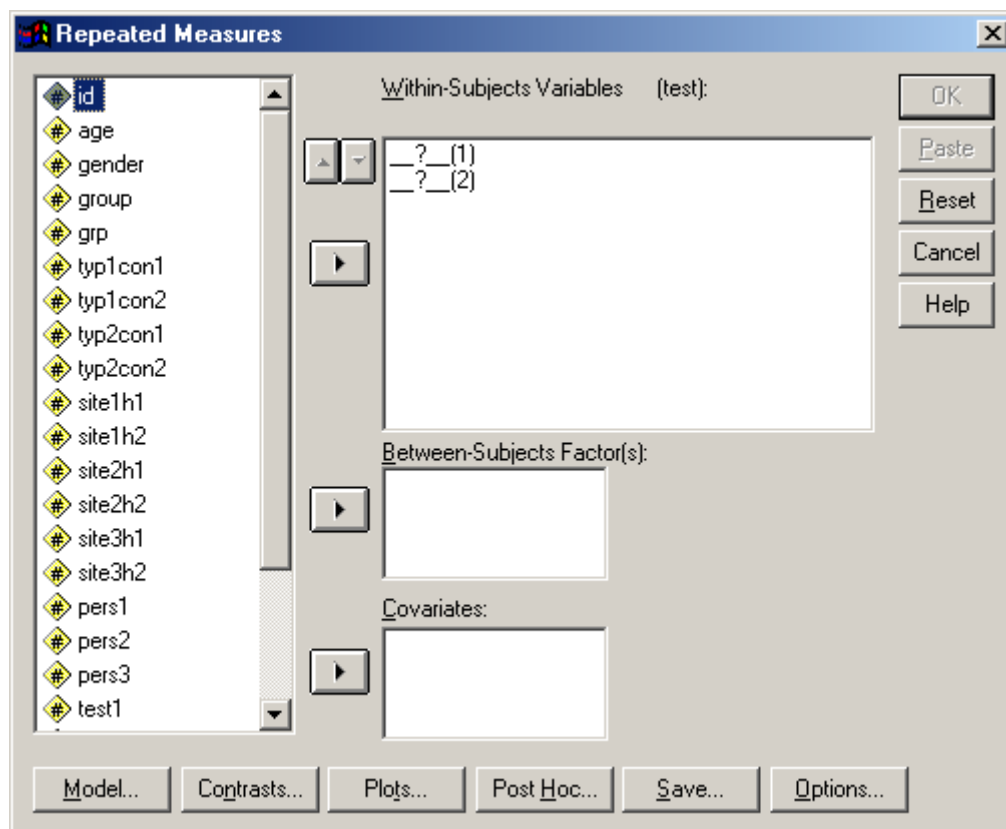
To carry out the analogous ANOVA, click on *Analyze* → *General Linear Model* → *Repeated Measures* and you will see this display:




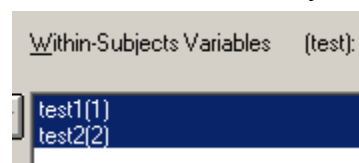
The first thing to enter is the *Within-Subject Factor Name*. This can be anything up to eight characters in length, as long as it doesn't duplicate a variable name which already exists in the dataset. Enter *test*. The second thing to enter is the *Number of Levels* in the within-subjects factor. Enter 2. Finally, click on the *Add* button, so that the display looks like this:



Now click on the *Define* button. The following display will appear:



GLM is waiting to have the names of the variables which make up the levels of the within-subject factor *type* specified in the *Within-Subjects Variables* window. You can either (a) click once on *test1* in the variable list at the left of the display, then click on the  button to transfer *test1* to the first slot in the *Within-Subjects Variables* window, then repeat the process with *test2*, or (b) click once on *test1*, hold down the *Ctrl* or *Shift* key, and click on *test2*, then on the arrow button. Either way, the *Within-Subjects Variables* window should look like this:



The numbers in brackets show the order in which the variables will be entered. There's nothing else that needs to be specified, but we could have a graph – click on the *Plots* button, and specify *test* as the *Horizontal Axis*.

#### Syntax

```
glm test1 test2/  
wsfactor test 2/  
plot=profile(test).
```

The repeated measures procedure produces a large amount of output, some of which is redundant for some analyses, such as the present one. For this example, we'll pick out just two tables to examine:

#### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	Sphericity Assumed	3.186	1	3.186	5.500	.021
	Greenhouse-Geisser	3.186	1.000	3.186	5.500	.021
	Huynh-Feldt	3.186	1.000	3.186	5.500	.021
	Lower-bound	3.186	1.000	3.186	5.500	.021
Error(TEST)	Sphericity Assumed	56.772	98	.579		
	Greenhouse-Geisser	56.772	98.000	.579		
	Huynh-Feldt	56.772	98.000	.579		
	Lower-bound	56.772	98.000	.579		

#### Tests of Between-Subjects Effects

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4452.006	1	4452.006	6388.629	.000
Error	68.293	98	.697		

The first table gives the result we're interested in and, as would be expected, the *p*-value is the same as that given by the paired *t*-test procedure. In fact the same *p*-value is given four times. That's because *GLM* checks whether a certain assumption has been met and provides results which are adjusted to the extent that the assumption isn't met. The assumption that *GLM* worries about will always be met when there are only two levels in the within-subject factor, so in this case the four results are identical. The reason the assumption is always met when there are only two levels in the within-subject factor will become evident in the next section.

The second table gives *Tests of Between-Subjects Effects*. Huh? There isn't a between-subject factor. What *GLM* is testing here is whether the sum of the two variables, *test1* and *test2* is different from zero. This isn't of any interest to us in the present case. (Whenever you see a ridiculously large *F*-ratio like 6388.6, you should be suspicious, and check what is actually being tested.)

The graph (not included here) shows that the mean for *test2* is higher than that for *test1*.

## 6.2 Three Levels in the Within-Subjects Factor

### 6.2.1 The Univariate and Multivariate Approaches to Repeated Measures

When the number of levels in a within-subject factor is three or greater, there are two different ways of carrying out the analysis. One is called the univariate approach to repeated measures. The other is the multivariate approach. (These methods are identical when there are only two levels.) *GLM* always prints out results for both methods, even when they are identical, so *GLM* users need to have some understanding of both the univariate and multivariate approach, and the assumptions underlying them, if they are to make sense of the mass of output produced by repeated measures analyses.

Although the univariate and multivariate methods are based on different methodologies, the difference that concerns most users is that the validity of the univariate approach depends on some rather restrictive assumptions about the data which are not necessary for the multivariate method. Therefore the multivariate results may be usable under circumstances in which the univariate results are questionable or at least need some adjustment. The next section considers the assumption which underlies the univariate approach.

#### 6.2.1.1 Circularity of the Variance-Covariance Matrix

The key assumption on which the univariate approach is based involves the variances and covariances of the variables which make up the levels of a within-subject factor. This assumption is analogous to the assumption of equal variances over groups in independent groups analyses and, as in that case, is needed if the estimate of the error variance, on which significance tests are based, is to be accurate. However, the presence of repeated measures makes the testing of the validity of the assumption more complicated than it is with independent groups. To make the discussion concrete, the variance-covariance matrix of *test1*, *test2* and *test3*<sup>5</sup> is shown below.

**Residual SSCP Matrix**

		TEST1 Test score 1	TEST2 Test score 2	TEST3 Test score 3
Covariance	TEST1 Test score 1	.6301	.0588	.0855
	TEST2 Test score 2	.0588	.6461	.1699
	TEST3 Test score 3	.0855	.1699	.7306
Correlation	TEST1 Test score 1	1.0000	.0921	.1260
	TEST2 Test score 2	.0921	1.0000	.2472
	TEST3 Test score 3	.1260	.2472	1.0000

Based on Type III Sum of Squares

From the matrix we can see, for example, that the variance of *test1* is .6301 and that its covariance with *test2* is .0588. (A correlation is simply a covariance standardised by dividing it by the product of the standard deviations of the two variables. Therefore, the

<sup>5</sup> The variance-covariance matrix can be obtained from *GLM* by clicking on the *Options* button and selecting *Residual SSCP Matrix*. In syntax, specify *print=rsscp*.

covariance of *test1* and *test2* is their correlation (.0921 in the bottom part of the table) multiplied by the standard deviations of both variables. Because the standard deviation is the square root of the variance, we can use the values from the top part of the table for the calculation. The covariance of *test1* and *test2* is  $.0921 * \sqrt{.6301} * \sqrt{.6461} = .0588$ .) For the record, the formula for the sample covariance is  $\Sigma \{[X_i - \text{mean}(X)][Y_i - \text{mean}(Y)]\} / n - 1$ .

The assumption on which the validity of the univariate approach to repeated measures is based is that the variance-covariance matrix has the property of *circularity*. Circularity exists when the sum of the variances of any two of the variables making up the levels of the within-subject factor, minus twice their covariance, is constant for all pairs of variables. That is, for example, that  $(.6301 + .6461 - 2 * .0588) = 1.159$  should equal  $(.6461 + .7306 - 2 * .1699) = 1.037$ , which isn't exactly true in this case.

Because the variance of the difference between two variables is the sum of their variances minus twice their covariance, namely  $\sigma_1^2 + \sigma_2^2 - 2\text{cov}_{12}$  (or  $\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2$ , where  $r$  is the correlation coefficient), this simply means that if the variance-covariance matrix is circular as defined above, the variances of the differences between the variables are the same for any pair of the variables (the variances of all of  $[test2 - test1]$ ,  $[test3 - test1]$  and  $[test2 - test3]$  in this case). It is this property that is usually mentioned when the requirements for a univariate analysis are discussed<sup>6</sup>. It is also often argued that circularity is unlikely to be present for many datasets which are analysed using repeated measures ANOVA. In order to explain how *GLM* tests for circularity, we need to make a detour. First we will consider the contrasts used in repeated measures and then how transformations of these contrasts enable *GLM* to carry out the tests for circularity which are shown in the output for repeated measures.

### 6.2.1.2 Contrasts in Repeated Measures

In the example we're going to look at (when we finally get there!), the data consist of three variables, *test1*, *test2* and *test3*. We're interested in whether there are significant differences between the means of these variables. One way of answering this question is to create a set of contrasts which summarise the variation among the means and to base the analysis on them. (This way of looking at the analysis is helpful for understanding both the univariate and multivariate approaches to repeated measures.) By default, *GLM* uses *polynomial* contrasts, which may be particularly appropriate for repeated measures over time, and are orthogonal (see the box below). (*Helmert* and *difference* contrasts, mentioned in Section 2.3.2, are also often used for within-subjects factors, and are also

---

<sup>6</sup> You will sometimes see reference to the assumption of *compound symmetry*, which exists when the variances of the variables making up the levels of the within-subject factor are all equal, and the covariances of the variables are also all equal. This assumption is more restrictive than circularity (which doesn't insist that the variances are all equal) and, if there is compound symmetry, there is also circularity. On the other hand, there may be circularity in the absence of compound symmetry. Most notably, this may occur when the within-subject factor has only two levels – the homogeneity of the treatment difference variances is ensured (or irrelevant) because there is only one difference, but the variances of the two variables may be different (Maxwell & Delaney, 1990).

**Box 3: Orthogonal Polynomial Contrasts**

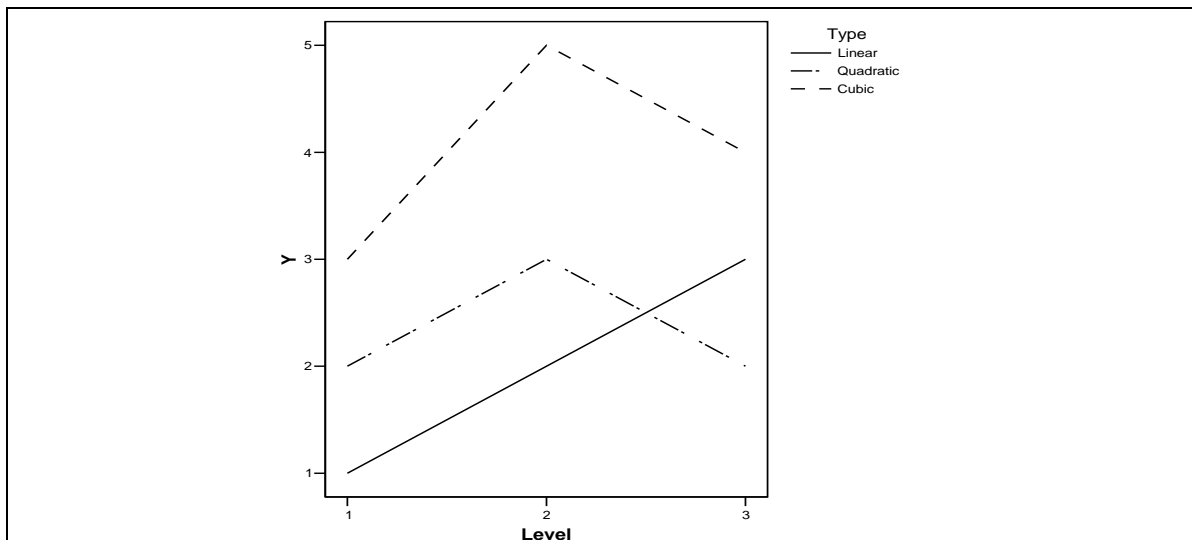
The difference between two means, 1 and 3, say, is obtained by subtracting one from the other:  $3 - 1 = 2$ , for example. In terms of contrast coefficients, this difference can be obtained by multiplying 3 by +1 and 1 by -1, i.e.,  $3 * 1 + 1 * -1 = 2$ . If the difference between two means is not zero, the result of the calculation ( $\text{Mean1} * 1 + \text{Mean2} * -1$ ) will not equal zero. (Whether the difference is significantly different from zero is a further question which *GLM* will answer for us.)

Polynomial contrast coefficients are constructed in such a way as to pick up, not merely differences between means of different treatment levels, but certain patterns of differences between means. As with any type of contrast, there can be as many polynomial contrasts as there are treatment levels, minus one. When there are three levels, such as *test1*, *test2* and *test3*, two polynomial contrasts are possible. The first is *linear* and the second is *quadratic*, meaning curvilinear with one 'bend'. (With four and five levels, there are *cubic* [curvilinear with two 'bends'] and then *quartic* [three 'bends'] contrasts.)

Means	Linear	Quadratic
	-1, 0, 1	1, -2, 1
1, 2, 3	$1 * -1 + 2 * 0 + 3 * 1 = 2$	$1 * 1 + 2 * -2 + 3 * 1 = 0$
2, 3, 2	$2 * -1 + 3 * 0 + 2 * 1 = 0$	$2 * 1 + 3 * -2 + 2 * 1 = -2$
3, 5, 4	$3 * -1 + 5 * 0 + 4 * 1 = 1$	$3 * 1 + 5 * -2 + 4 * 1 = -3$

The coefficients for linear and quadratic contrasts are shown at the top of the second and third columns of the above table. To illustrate how they work, we'll apply them to three sets of means. The means are shown in the first column of the table and also in the graph below. The first set of means, 1, 2 and 3, shows a purely linear increase. The second set of means, 2, 3 and 2, shows no linear trend but do go up, and then down to the same level, a simple curvilinear pattern. The pattern of the third set of means, 3, 5 and 4, is more complicated: there is both a slight upward linear trend (the second and third means, 5 and 4, are both higher than the first) and a curvilinear trend, going from 3 to 5, then dropping again, but not right back to the first level.





In the second and third columns of the table, the means are multiplied by the linear and quadratic contrast coefficients. It can be seen that, for the linearly-increasing means, 1, 2 and 3, the result of the linear contrast is 2, while the result of the quadratic contrast is zero. On the other hand, for the means with the curved pattern, 2, 3 and 2, the result of the linear contrast is zero while the result of the quadratic contrast is -2 (the sign is just a consequence of the signs of 1, -2 and 1: the coefficients could equally well have been -1, 2 and -1, in which case the result would have been positive). For the 'mixed' pattern of means, the results of both contrasts are non-zero. It is in this way that polynomial contrasts pick up the various possible components of trend in a set of means.

Finally, the orthogonality of the polynomial contrasts can be demonstrated as follows: (a) In both cases, the sum of the coefficients is zero:  $-1 + 0 + 1 = 0$ , and  $1 - 2 + 1 = 0$ . (b) The sum of the products of the corresponding coefficients in the two contrasts is zero:  $(-1 * 1) + (0 * -2) + (1 * 1) = 0$ .

orthogonal.) Two contrasts are needed to summarise the information in three measures, such as *test1*, *test2* and *test3*. The test of whether the means of these three variables are different can be seen as a test of whether the results of the two contrasts are zero. The difference between the univariate and multivariate approaches to repeated measures lies in the way this test is carried out. But before we can deal with the actual analyses, we need to return to the question of how GLM tests whether the key assumption for the univariate approach to repeated measures is met.

### 6.2.1.3 Testing Circularity by Testing Sphericity

In Section 6.2.1.1, it was noted that the assumption on which the univariate method for analysing repeated measures depends is met if the variance-covariance matrix of the variables making up the levels of the within-subjects factor has the property of circularity. How does *GLM* test for this property? In a nutshell, the circularity of a variance-covariance matrix can be checked by testing the *sphericity* of the result of multiplying the variance-covariance matrix by a set of contrast coefficients which have been *orthonormalised*.

## 6.2.1.4 Orthonormalisation

We'll deal with the second term first. The *ortho*- part simply refers to orthogonality, meaning that the contrasts have to be orthogonal, like the orthogonal polynomials described in the previous box and the Helmert and difference contrasts also mentioned earlier. The second part means that the contrast coefficients have been *normalised* (the reason for this is explained below). They are normalised by dividing each by the 'norm'

Contrasts	Type	Original coefficients	Norm (length)	Orthonormalised coefficients
Polynomials	Linear	-1, 0, 1	$\sqrt{(-1^2 + 0^2 + 1^2)} = \sqrt{2} = 1.4142$	$-1/\sqrt{2}, 0/\sqrt{2}, 1/\sqrt{2} =$ <b>-.7071, 0, .7071</b>
	Quadratic	1, -2, 1	$\sqrt{(1^2 - 2^2 + 1^2)} = \sqrt{6} = 2.4495$	$1/\sqrt{6}, -2/\sqrt{6}, 1/\sqrt{6} =$ <b>.4082, -.8165, .4082</b>
Helmert	1 vs (2 + 3)/2	1, -.5, -.5	$\sqrt{(1^2 - .5^2 - .5^2)} = \sqrt{1.5} = 1.2247$	$1/\sqrt{1.5}, -.5/\sqrt{1.5}, -.5/\sqrt{1.5} =$ <b>.8165, -.4082, -.4082</b>
	2 vs 3	0, 1, -1	$\sqrt{(1^2 - 1^2)} = \sqrt{2} = 1.4142$	$0/\sqrt{2}, 1/\sqrt{2}, -1/\sqrt{2} =$ <b>0, .7071, -.7071</b>

or 'length' of the set of coefficients. This operation is shown in the above table for two polynomial and two Helmert contrasts. The norm or length of a set of contrast coefficients is the square root of the sum of their squares. For example, the norm of the linear contrasts [-1, 0, 1] is  $\sqrt{(-1^2 + 0^2 + 1^2)} = \sqrt{2} = 1.4142$ . The last column of the table shows the orthonormalised coefficients produced by dividing each of the original coefficients by the norm.

It can be seen that the relationships between the original coefficients are preserved by the orthonormalised coefficients; considering the quadratic contrast coefficients, for example,  $.4082/-.8165 = 1/-2$ . However, the orthonormal coefficients have a property which the original coefficients do not, in that they are of unit length: the sum of the squared orthonormal coefficients is always one. Because of this property, the application of any set of orthonormal coefficients to variables which have the same variance will result in variables which also have equal variance. To illustrate this, consider the randomly-

Descriptive Statistics

	N	Std. Deviation	Variance
A	99	2.17010	4.709
B	99	2.15098	4.627
C	99	2.15794	4.657
LIN	99	2.14650	4.607
QUAD	99	3.71861	13.828
NLIN	99	1.51781	2.304
NQUAD	99	1.51811	2.305

generated variables  $a$ ,  $b$  and  $c$ , the standard deviations and variances of which, shown in the first three lines of the table above, are very similar<sup>7</sup>. The lines in the table labelled *lin* and *quad* show the results of multiplying  $a$ ,  $b$  and  $c$  by the polynomial contrast coefficients  $[-1, 0, 1]$  and  $[1, -2, 1]$  respectively. The variances of the two results are very different (4.61 and 13.83 respectively). The lines in the table labelled *nlin* and *nquad* respectively are the result of multiplying  $a$ ,  $b$  and  $c$  by the orthonormalised versions of the coefficients, namely  $[-.7071, 0, .7071]$  and  $[.4082, -.8165, .4082]$  respectively. The variances of the two results are virtually identical (2.304 and 2.305).

Finally, though, we get to the property of the orthonormalised contrast coefficients which provides the basis for the test of circularity. This property will be illustrated with an example.

Let Matrix 1 be the matrix of the orthonormalised polynomial contrast coefficients:

-.7071	0	.7071
.4082	-.8165	.4082

Let Matrix 2 be the variance-covariance matrix for the variables  $a$ ,  $b$  and  $c$ :

	$a$	$b$	$c$
$a$	4.7094	2.3975	2.3793
$b$	2.3975	4.6267	2.3034
$c$	2.3793	2.3034	4.6567

Let Matrix 3 be the transpose of Matrix 1 (i.e., Matrix 1 is 'flipped' so rows become columns and columns become rows):

-.7071	.4082
0	-.8165
.7071	.4082

The crucial property is that if Matrix 2 (the variance-covariance matrix of the variables defining the within-subject factor) is circular, then

Matrix1 \* Matrix 2 \* Matrix3 = Matrix4, which has the property of sphericity.

This is the test of circularity used by *GLM*: If the variance-covariance matrix of the variables making up the within-subject factor has the property of circularity, then the pre-multiplication of that matrix by a matrix of orthonormalised contrasts, and the post-multiplication by the transpose of the same matrix of orthonormalised contrasts, will give rise to a matrix which has the property of sphericity. This matrix (Matrix4 in this example) is the variance-covariance matrix of the variables which are produced by multiplying the original variables by the orthonormalised contrast coefficients. To make this concrete, imagine that we used the following commands to create the orthonormalised contrasts:

---

<sup>7</sup>  $a$ ,  $b$  and  $c$  were calculated with the properties necessary to meet the requirements of the univariate approach to repeated measures. The commands used to compute them are given in Appendix 1.

$compute\ nlin = a * -.7071 + b * 0 + c * .7071.$   
 $compute\ nquad = a * .4082 + b * -.8165 + c * .4082.$

then obtained the variance-covariance matrix for the two variables *nlin* and *nquad*. In this example, the matrix is:

**Variance-Covariance Matrix**

It is this matrix which is tested for sphericity.

		NLIN	NQUAD
Covariance	NLIN	2.3037	.0392
	NQUAD	.0392	2.3047

We can summarise what we've covered so far by this rule:

*If the variance-covariance matrix of the transformed variables (transformed by multiplication by orthonormalised contrast coefficients) is spherical, then the variance-covariance matrix of the original variables is circular, and the univariate approach to repeated measures is valid.*

In the next section we will define sphericity, and see if the product of the above multiplication has this property.

#### 6.2.1.5 Sphericity

The definition of a spherical matrix is simple – it has a constant on the diagonal and zeroes elsewhere. For example, the two following matrices are spherical:

3.5	0	0
0	3.5	0
0	0	3.5

1	0	0
0	1	0
0	0	1

(In addition to being spherical, the second is an *identity* matrix.)

The following matrices are not spherical:

3.5	1.1	2.2
1.1	3.5	0
2.2	0	3.5

3	0	0
0	4	0
0	0	5

Now let's consider the result of the multiplication Matrix1 \* Matrix2 \* Matrix3 above (the full calculation is shown in Appendix 3):

2.3037	0.0391
0.0391	2.3037

This isn't exactly spherical, but it seems close. Is it or isn't it? To answer this question, *GLM* uses a test described by Mauchly (1940). The calculation of the Mauchly's test statistic,  $W$ , and the method used to test its significance, is described in Winer, Brown & Michels (1991), pp 255-259 (this book also has a detailed discussion of the assumptions underlying the univariate approach to repeated measures).  $W$  ranges from 1 (sphericity) to 0. The smaller the value of  $W$ , the greater the departure from sphericity. For the record,  $W$  is 1.00 for this matrix, so it would be appropriate to analyse the variables  $a$ ,  $b$  and  $c$  with the univariate approach to repeated measures.

So is that it? Use the univariate approach to repeated measures unless the value of  $W$  calculated by *GLM* shows a significant departure from 1.00? Not quite. *GLM* also reports results which show the extent of the departure of the data from sphericity (and therefore from circularity) and which can be used to adjust the significance tests in a way which makes them more conservative (less likely to reach a given significance level) the greater the departure from circularity. These will be described in the next section.

#### 6.2.1.6 Epsilon -- The Extent of Departure from Circularity


Epsilon ( $\varepsilon$ ) is a measure of the extent to which a variance-covariance matrix departs from circularity. As with  $W$ ,  $\varepsilon = 1$  when a matrix is perfectly circular, and becomes smaller than one with departures from circularity. Several writers have suggested ways of calculating  $\varepsilon$  from an obtained variance-covariance matrix, and using  $\varepsilon$  to adjust the degrees of freedom of significance tests of effects involving within-subjects factors. Departures from circularity tend to make significance tests too liberal (i.e., more likely to reject the null hypothesis, when it is in fact true, than the nominal alpha would suggest), so the adjustments with  $\varepsilon$  are used to make the tests more conservative. We'll see that, along with the unadjusted significance tests for within-subject factors, *GLM* gives three adjusted results. The first is due to Greenhouse and Geisser (1959) and tends to be too conservative, the second is due to Huynh and Feldt (1976) and tends to be too liberal. The third is the 'bottom line', the most conservative possible result in terms of the departure from circularity. The  $\varepsilon$  values, and the adjustments based on them, will be discussed again when we look at some SPSS output. Details of the calculations of  $\varepsilon$  are given in Winer, Brown & Michels (1991), pp 251-255, and in Stevens (1992), pp 447-449 and pp 492-493.

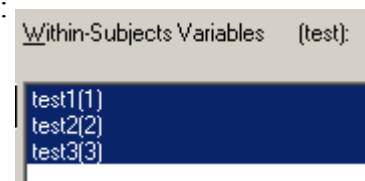
#### 6.2.1.7 The Multivariate Approach to Repeated Measures

Multivariate analysis is covered in another handbook, *Multivariate Analyses with manova and GLM*, so if you would like a deeper understanding at this point, it should help if you read the other handbook before continuing here. The main point to make at this stage is that the multivariate approach does not depend on the circularity assumption required for the univariate approach. In the multivariate approach, the  $k - 1$  contrasts which have been created to summarise the  $k$  means for the  $k$  variables representing the levels of the within-subject factor are simply treated as a set of dependent variables on which a multivariate analysis is carried out. In a nutshell, multivariate analysis involves combining a set of dependent variables in an optimal way for a given analysis. When there is a single within-subject variable, the question is whether, when they are considered together, the

dependent variables (which are contrasts) are different from zero. In the multivariate approach to repeated measures, the two contrasts are weighted appropriately, combined, then a test is carried out to see if the composite variable is different from zero. This does not require any assumptions about the variances or covariances of the variables. The results of the univariate and multivariate approaches will now be considered for an actual analysis.

### 6.2.2 Carrying out a One-way Repeated Measures ANOVA with Three Levels in the Within-Subject Factor

As for the two-level analysis in Section 6.1, click on *Analyze* → *General Linear Model* → *Repeated Measures*. Enter *test* as the *Within-Subject Factor Name*, and 3 as the *Number of Levels* and click on *Add*. Click on the *Define* button, and select the variables *test1*, *test2* and *test3* and click on the  button so that the the three variables appear in the *Within-Subjects Variables* window like this:



It's always good to have a graph, so click on the *Plots* button and ask for a graph with *test* on the *x-axis* (don't forget to click on the *Add* button). Finally, for the purposes of this demonstration, click on the *Options* button and check the *Residual SSCP matrix* box (*SSCP* stands for *Sums of Squares and Cross-Products*) and also the *Transformation matrix* box. These are not things you would ask for routinely.

#### Syntax

```
glm test1 test2 test3/
  wsfactor test 3/
  print=rsscp test(mmatrix)/
  plot=profile(test).
```

Specifying *test(mmatrix)* in the syntax is equivalent to checking the *Transformation matrix* box in point-and-click. The word *test* would be used before (*mmatrix*) even if it wasn't the name of the within-subject factor.

## General Linear Model

### Within-Subjects Factors

Measure: MEASURE\_1

TEST	Dependent Variable
1	TEST1
2	TEST2
3	TEST3

The first table produced by *GLM* shows the name of the within-subject factor, the variables which define its levels, and the order in which they are entered. In designs with more than one within-subject factor, this table allows you to check that *GLM* has interpreted your commands as you intended.

**Bartlett's Test of Sphericity<sup>a</sup>**

Likelihood Ratio	.012
Approx. Chi-Square	8.653
df	5
Sig.	.124

Tests the null hypothesis that the residual covariance matrix is proportional to an identity matrix.

a.

Design: Intercept

Within Subjects Design: TEST

has been multiplied by the orthonormalised contrast coefficients. Therefore it isn't a test of circularity, but simply tells us whether the variance-covariance matrix of the variables (*test1*, *test2* and *test3* in this case) departs from one in which all the variables have the same variance and are uncorrelated. The non-significant result in this example suggests that the data don't depart significantly from this pattern. This would normally be surprising, because measures taken from the same subjects tend to be correlated (this is

**Bartlett's Test of Sphericity<sup>a</sup>**

Likelihood Ratio	.000
Approx. Chi-Square	68.847
df	5
Sig.	.000

Tests the null hypothesis that the residual covariance matrix is proportional to an identity matrix.

a.

Design: Intercept

Within Subjects Design: LETTER

*Bartlett's Test of Sphericity?* We aren't prepared for this! Relax. *GLM* only provides this table when asked for the *SSCP* matrix, and Bartlett's test is different from the one that we discussed at some length above. The present test is performed on the actual variance-covariance matrix of the variables which represent the levels of the within-subject factor, not one that has been multiplied by the orthonormalised contrast coefficients. Therefore it isn't a test of circularity, but simply tells us whether the variance-covariance matrix of the variables (*test1*, *test2* and *test3* in this case) departs from one in which all the variables have the same variance and are uncorrelated. The non-significant result in this example suggests that the data don't depart significantly from this pattern. This would normally be surprising, because measures taken from the same subjects tend to be correlated (this is one of the reasons for using repeated measures designs), but in this case the data are partly artificial, and correlations weren't built into them. It is worth noting that the result of the test for variables *a*, *b* and *c* (shown at left), which were generated to have similar variances, and correlations of around 0.5 with each other, is highly significant.

results for the multivariate approach. The meaning

The next table (below) gives the

**Multivariate Tests<sup>b</sup>**

Effect	Value	F	Hypothesis df	Error df	Sig.
TEST Pillai's Trace	.235	14.938 <sup>a</sup>	2.000	97.000	.000
Wilks' Lambda	.765	14.938 <sup>a</sup>	2.000	97.000	.000
Hotelling's Trace	.308	14.938 <sup>a</sup>	2.000	97.000	.000
Roy's Largest Root	.308	14.938 <sup>a</sup>	2.000	97.000	.000

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: TEST

of the four different multivariate statistics are discussed in detail in the handbook on multivariate analysis. Here it's enough to note that all the statistics give the same answer. Probably the most commonly-quoted multivariate statistic is Wilks' Lambda. If you wanted to refer to this result, you would write something like 'The overall test effect was significant: Wilks' Lambda = .77,  $F(2,97) = 14.94$ ,  $p < .0005$ '. The distributions of the multivariate statistics are complex, and *GLM* generally uses an approximation based on the *F*-distribution to test the significance of the multivariate results. However, in this case the significance test is based on the exact distribution, and the output notes this. The

degrees of freedom given in the table are for the distribution of the multivariate statistic and, while they are affected by the number of levels and cases, they generally won't be the same as those for the univariate approach.

The table below is one we've put a lot of time into preparing to interpret: it shows the results of the Mauchly test of sphericity (which tells us whether the variance-covariance matrix departs from circularity) and also the  $\varepsilon$  values, which tell us the extent of the departure from circularity. Unfortunately (but fortunately for the analysis), the contents of the table aren't a great test of our mettle: Mauchly's  $W$  is almost one, and is clearly not significant, and the Greenhouse-Geisser and Huynh-Feldt  $\varepsilon$  values are 1.00 or close to it. Therefore, the circularity assumption is met, and there's no reason why we shouldn't accept the univariate results at face value.

**Mauchly's Test of Sphericity<sup>b</sup>**

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TEST	.993	.663	2	.718	.993	1.000	.500

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

a. 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.

b.

Design: Intercept

Within Subjects Design: TEST

The combined univariate results are shown in the next table. As we would expect, the

**Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	Sphericity Assumed	15.681	2	7.840	13.896	.000
	Greenhouse-Geisser	15.681	1.986	7.894	13.896	.000
	Huynh-Feldt	15.681	2.000	7.840	13.896	.000
	Lower-bound	15.681	1.000	15.681	13.896	.000
Error(TEST)	Sphericity Assumed	110.587	196	.564		
	Greenhouse-Geisser	110.587	194.674	.568		
	Huynh-Feldt	110.587	196.000	.564		
	Lower-bound	110.587	98.000	1.128		

basic result ('Sphericity Assumed') is highly significant, and the adjusted results (even the 'Lower-bound' results, which assume the greatest possible divergence from circularity) are also highly significant.<sup>8</sup>

<sup>8</sup> It's not possible to tell this from the table without showing more decimal places, but there is a marked (though unimportant) difference between the  $p$ -values for the 'Sphericity Assumed' result [ $F(2, 196) = 13.896, p = .000002272$ ] and the 'Lower-bound' result [ $F(1, 98) = 13.896, p = .000323$ ]. This shows that the lower-bound result is indeed conservative.



By default, the next table in the *GLM* output shows the result for each of the two polynomial contrasts, linear and quadratic. These are the two univariate results which are combined to make up the overall result in the univariate approach to repeated measures. As can be seen, both the linear and quadratic components are significant, and this will make sense when we see the plot later in the output.

It's worth noting that, when orthonormalised contrast coefficients are used, the sums of squares of the univariate tests add up exactly to the sums of squares shown in the above table:  $4.703 + 10.977 = 15.681$ , and  $58.293 + 52.294 = 110.587$ . When contrasts

#### Tests of Within-Subjects Contrasts

Measure: MEASURE\_1

Source	TEST	Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	Linear	4.703	1	4.703	7.907	.006
	Quadratic	10.977	1	10.977	20.572	.000
Error(TEST)	Linear	58.293	98	.595		
	Quadratic	52.294	98	.534		

are orthonormalised, they are independent and equally-weighted, which gives the resulting sums of squares the property of adding up to the total sums of squares; if the contrasts are not orthonormalised, the sums of squares corresponding to each contrast will no necessarily add up to total sum of squares.

#### Tests of Between-Subjects Effects

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	6275.753	1	6275.753	7144.781	.000
Error	86.080	98	.878		

As with the equivalent table in the two-level example discussed in Section 6.1, the table showing the between-subject effects isn't of great interest, as it just tells us that the sum of *test1*, *test2* and *test3* is significantly different from zero.

And still the output keeps coming. The tables below, produced because we asked to see the transformation matrix (point-and-click) or the *test(mmatrix)* (syntax), show the contrast coefficients used in the analysis.

### Transformation Coefficients (M Matrix)

#### Average

Measure: MEASURE\_1

Transformed Variable: AVERAGE

TEST1	Test score 1	.5774
TEST2	Test score 2	.5774
TEST3	Test score 3	.5774

The first table shows that even the coefficients used to produce the sum of *test1*, *test2* and *test3*, used in the between-subject part of the analysis were normalised:  $.5774^2 + .5774^2 + .5774^2 = 1$ . This means that the

original coefficients were 1, 1 and 1, and were divided by  $\sqrt{(1^2 + 1^2 + 1^2)} = \sqrt{3} = 1.7321$ .

### TEST<sup>a</sup>

Measure: MEASURE\_1

Dependent Variable	TEST	
	Linear	Quadratic
TEST1 Test score 1	-.7071	.4082
TEST2 Test score 2	.0000	-.8165
TEST3 Test score 3	.7071	.4082

The second table shows the coefficients for the linear and quadratic contrasts, which have the values we would expect from previous discussions.

- a. The contrasts for the within subjects factors are:

TEST: Polynomial contrast

The final table, produced because we asked for the RSSCP matrix, shows the variance-covariance matrix for the

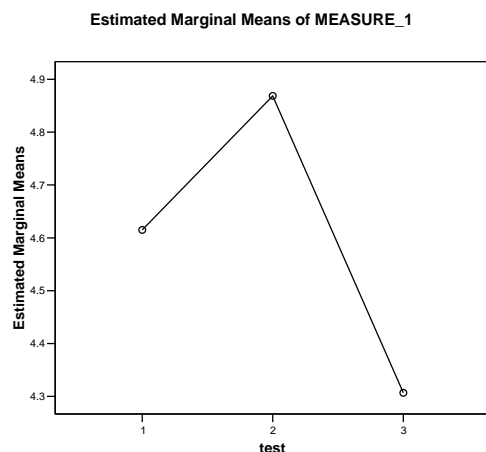
### Residual SSCP Matrix

		TEST1 Test score 1	TEST2 Test score 2	TEST3 Test score 3
Sum-of-Squares and Cross-Products	TEST1 Test score 1	61.746	5.760	8.381
	TEST2 Test score 2	5.760	63.319	16.645
	TEST3 Test score 3	8.381	16.645	71.603
Covariance	TEST1 Test score 1	.630	.059	.086
	TEST2 Test score 2	.059	.646	.170
	TEST3 Test score 3	.086	.170	.731
Correlation	TEST1 Test score 1	1.000	.092	.126
	TEST2 Test score 2	.092	1.000	.247
	TEST3 Test score 3	.126	.247	1.000

Based on Type III Sum of Squares

three variables, the SSCP matrix from which the variances and covariances can be derived, and the correlations (standardised covariances) between the variables. The sum of squares for a given variable is the sum of the squared deviations of each subject's score from the overall mean for that variable. The variance of the variable is then the sum of squares divided by the number of cases minus one. For example, for *test1*,  $.630 = 61.746/98$ . The cross product of *test1* and *test2* for each subject (for example) is the product (subject's score on *test1* – overall mean of *test1*) \* (subject's score on *test2* – overall mean of *test2*). This product is summed over subjects and divided by the number of subjects minus one to give the covariance.

As can be seen from the table, the correlations between the variables are not very great, which accounts for the non-significance of the Bartlett test of sphericity discussed earlier.



Finally, we get the profile plot, which shows the means of *test1*, *test2* and *test3*. This plot shows why the linear and quadratic contrasts were significant: there is a definite curve, as well as an overall downward trend.

### 6.2.2.1 Follow-up Comparisons of Means

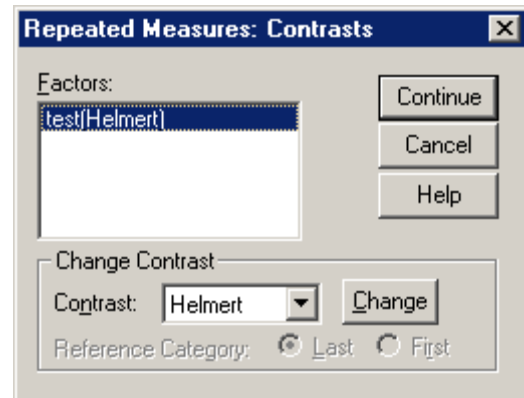
In the present example, the polynomial linear and quadratic contrasts may be sufficient for the purposes of the analysis. However, for some purposes, pairwise or other comparisons may be needed. With repeated measures, comparisons may be carried out (i) by specifying contrasts [the syntax for this method is different from that used for between-subject factors], (ii) by asking for estimated marginal means and (iii) by using the *mmatrix* subcommand (syntax only), which is the repeated measures equivalent of the *lmatrix* subcommand described in Section 4.6. *Post hoc* comparisons are not available for within-subject factors in the multivariate layout. There is also little point in asking for parameter estimates with repeated measures, because they only tell us whether the mean of each of the variables making up the levels of the within-subject factor differs significantly from zero, which is not something there is usually much interest in.

For the purposes of demonstration, we'll carry out a follow-up analysis in which we use all three of these methods at once, and test two different sets of contrasts, something which might not happen very often in practice. To make things a bit more sensible, and also concrete, we'll identify each set of contrasts with a hypothetical researcher. As Researcher 1, we want to test just two *a priori* contrasts, *test1* versus the mean of *test2* and *test3*, and *test2* versus *test3*. These tests can be done with the *Contrast* subcommand (because they are the standard Helmert contrast) or the *mmatrix* subcommand (in practice we wouldn't need to use both, because they'll give identical results). Because these contrasts are *a priori* (i.e., we decided we wanted to test them before we laid our hands on the results), we'll adjust the *p*-values for just the two contrasts, rather than for all the contrasts we might have done, and use a Bonferroni-adjusted value of  $.05/2 = .025$ .

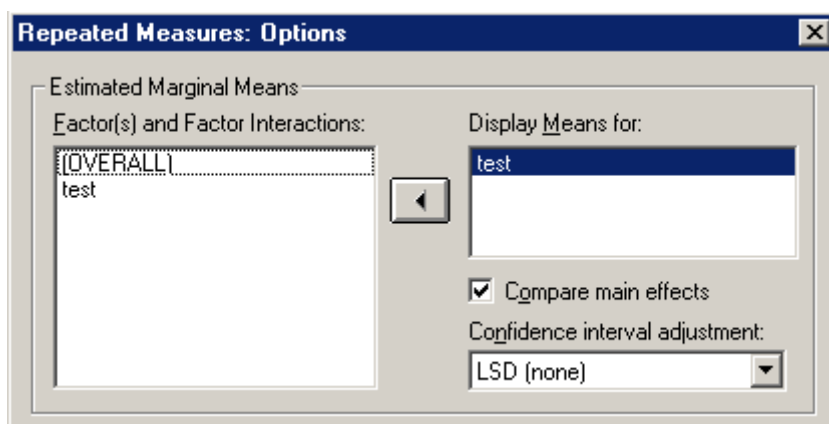
As Researcher 2, we want to make all possible pairwise comparisons of the three variables *test1*, *test2* and *test3*, and because we haven't any particular hypothesis about the possible differences, we'll consider them to be *post hoc*. We'll imagine that we decided ahead of time that we would limit ourselves to pairwise comparisons (we might have said, 'It just wouldn't make any sense to compare the average of *test1* and *test2* with *test3*', for example), so we don't have to make an adjustment which covers all possible comparisons involving the three variables, including complex ones [such as  $test2 - \text{mean}(test1, test3)$ ]. (If we hadn't wanted to make even that decision ahead of time ['I'll wait and see what the results look like – I might want to compare the average of *test1* and *test2* with *test3*', for example], then we should use the Scheffé method, which allows for any *post hoc* comparison we could think of.)

Because, as Researcher 2, we didn't commit ourselves ahead of time, regardless of how the results came out, to a particular one or two pairwise comparisons, we need to cover all the pairwise tests we could have done, regardless of how many we actually end up doing; for example, it might be very obvious, once we look at the results, that the means for *test1* and *test2* are virtually identical, so there would be no point in making a formal comparison; however, we could have wanted to make such a comparison, so we must allow for it. For the pairwise comparisons, we'll use the *emmeans* subcommand, and adopt a Bonferroni-adjusted *p*-value of  $.05/3 = .0167$ .

First, go through the steps described on Section 6.2.2 to set up the analysis, then click on the *Contrasts* button. Select *Helmert* as the contrast for *test*, and click on the *Change* button. The display should look like this: Click on *Continue*.



Then, to have *GLM* compare estimated marginal means, click on the *Options* button, and select *test* and check the *Compare main effects* box, so that the display looks like this:



While on the *Options* display, check the following box:



### Syntax

```
glm test1 test2 test3/
wsfactor test 3 helmert/
emmeans=table(test) compare(test)/
mmatrix="t1vt23" test1 2 test2 -1 test3 -1; "t2vt3" test1 0 test2 1 test3 -1/
print=test(mmatrix).
```

**Notes:** (a) For within-subject factors, contrasts are specified in the *wsfactor* subcommand, rather than in a separate *contrast* subcommand.  
 (b) The syntax of the *mmatrix* subcommand is different from that of the *lmatrix* subcommand in that a label can occur for each contrast in a set of contrasts (not just for each set of contrasts) and the dependent variables, rather than effect names, are specified. Each variable is followed by a contrast coefficient.  
 (c) "t1vt23", "t2vt3", etc are labels only – they have no effect at all on the way the contrasts are carried out. "t1vt23" is short for *test1 versus the mean of test2 and test3*.

The following are the parts of the output showing the results of three contrasts. The first

## Tests of Within-Subjects Contrasts

Measure: MEASURE\_1

Source	TEST	Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	Level 1 vs. Later	.074	1	.074	.080	.777
	Level 2 vs. Level 3	31.263	1	31.263	30.146	.000
Error(TEST)	Level 1 vs. Later	89.658	98	.915		
	Level 2 vs. Level 3	101.631	98	1.037		

table shows the result of the Helmert contrasts. Notice that the two contrasts are nicely labelled (in the *TEST* column). The table shows that there was no significant difference between the mean of *test1* and the combined means of *test2* and *test3* (.777 is clearly greater than .025). There was a clear difference between *test2* and *test3* (.000, < .0005, is clearly less than .025). An important point to notice is that *GLM* has used a different error term for each contrast, rather than using the pooled error term (this was also the case with the polynomial contrasts shown earlier). This strategy avoids the possible problems of lack of circularity in the variance-covariance matrix.

The following tables show the coefficients which were used for the Helmert contrasts:

## Transformation Coefficients (M Matrix)

## Average

Measure: MEASURE\_1

Transformed Variable: AVERAGE

TEST1 Test score 1	.333
TEST2 Test score 2	.333
TEST3 Test score 3	.333

Note that neither the contrast coefficients for the average of the three tests nor those for the Helmert contrasts are normalised, although the Helmert contrasts are orthogonal.

TEST<sup>a</sup>

Measure: MEASURE\_1

Dependent Variable	TEST	
	Level 1 vs. Later	Level 2 vs. Level 3
TEST1 Test score 1	1.000	.000
TEST2 Test score 2	-.500	1.000
TEST3 Test score 3	-.500	-1.000

a. The contrasts for the within subjects factors are:

TEST: Helmert contrast

## Custom Hypothesis Tests

## Transformation Coefficients (M Matrix)

Dependent Variable	Transformed Variable	
	t2vt1	t3vt12
TEST1 Test score 1	-1.000	-1.000
TEST2 Test score 2	1.000	-1.000
TEST3 Test score 3	.000	2.000

The next table shows the coefficients used for the contrasts specified in the *mmatrix* subcommand (syntax only). Notice that the coefficients have not been normalised, and that in this and the next table, *GLM* has used the labels we assigned in the subcommand.

The results of the two contrasts are shown in the next table. The *p*-values are identical to those shown earlier, as expected. The table gives information about the size and direction of the differences (e.g., the mean for *test3* is .562 smaller than that for *test2*) which makes it more informative than the table produced by the *contrast* subcommand.

**Contrast Results (K Matrix)**

Contrast <sup>a</sup>		Transformed Variable	
		t1vt23	t2vt3
L1	Contrast Estimate	.055	.562
	Hypothesized Value	0	0
	Difference (Estimate - Hypothesized)	.055	.562
	Std. Error	.192	.102
	Sig.	.777	.000
	95% Confidence Interval for Difference	Lower Bound -.327	.359
		Upper Bound .436	.765

a. Estimable Function for Intercept

Because *GLM* will have used separate error terms for each contrast, each with 98 degrees of freedom, these two results could be reported as  $t(98) = .055/.192 = .29$ ,  $p = .777$  and  $t(98) = .562/.102 = 5.51$ ,  $p < .0005$ . (Note that, due to rounding error, these  $t$ -values differ slightly from the square root of the  $F$ -values shown in the earlier *contrast* table.)

Finally, the results of the pairwise comparisons of the estimated marginal means are given:

## Estimated Marginal Means TEST

**Transformation Coefficients (M Matrix)**

Measure: MEASURE\_1

Dependent Variable	TEST		
	1	2	3
TEST1 Test score 1	1	0	0
TEST2 Test score 2	0	1	0
TEST3 Test score 3	0	0	1

The coefficient matrix for the pairwise tests isn't very informative, merely showing that the tests involved each variable weighted equally.

**Estimates**

Measure: MEASURE\_1

TEST	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	4.615	.080	4.457	4.773
2	4.869	.081	4.708	5.029
3	4.307	.086	4.136	4.477

The next table shows the means which were compared.

**Pairwise Comparisons**

Measure: MEASURE\_1

(I) TEST	(J) TEST	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	-.254*	.108	.021	-.468	-3.902E-02
	3	.308*	.110	.006	9.070E-02	.526
2	1	.254*	.108	.021	3.902E-02	.468
	3	.562*	.102	.000	.359	.765
3	1	-.308*	.110	.006	-.526	-9.070E-02
	2	-.562*	.102	.000	-.765	-.359

Based on estimated marginal means

\*. The mean difference is significant at the .050 level.

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

The above table shows the results of the comparisons between each pair of means. The differences between *test3* and each of the other two variables are significant ( $p < .0167$ ), but the difference between *test1* and *test2* is not ( $.021 > .0167$ ).





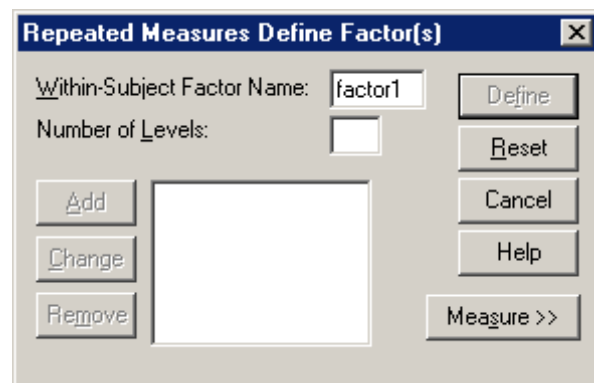
## 7. Two Within-Subjects Factors

When there are two or more within-subjects factors, no new complications are introduced, but a little care is needed when specifying the factor names and then specifying the variables in an order which corresponds to the order in which the factors have been named. For this example, we'll use fictional data which arise from a study in which six measurements were made from each subject's scalp – three on each hemisphere. The variables are in this order in the SPSS dataset:

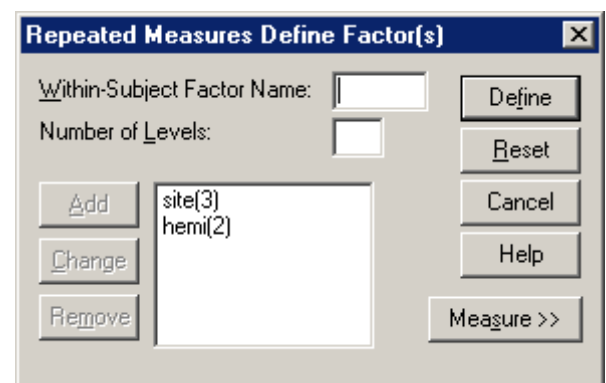
```
SITE1H1
SITE1H2
SITE2H1
SITE2H2
SITE3H1
SITE3H2
```

The two measurements for Site 1 come first -- Hemisphere 1 then Hemisphere 2 -- then the two measurements for Site 2, and so on. Notice that the levels of the Hemisphere factor change most quickly. Considering the first two variables, for example, Site remains at 1 while Hemisphere changes from 1 to 2. The rule used by SPSS is that the levels of later-named factors are expected to change more quickly than those of earlier-named factors. So, if we wanted to enter the variables into the analysis in the above order, Site would be named first, then Hemisphere. This will become clear as we go through the example.

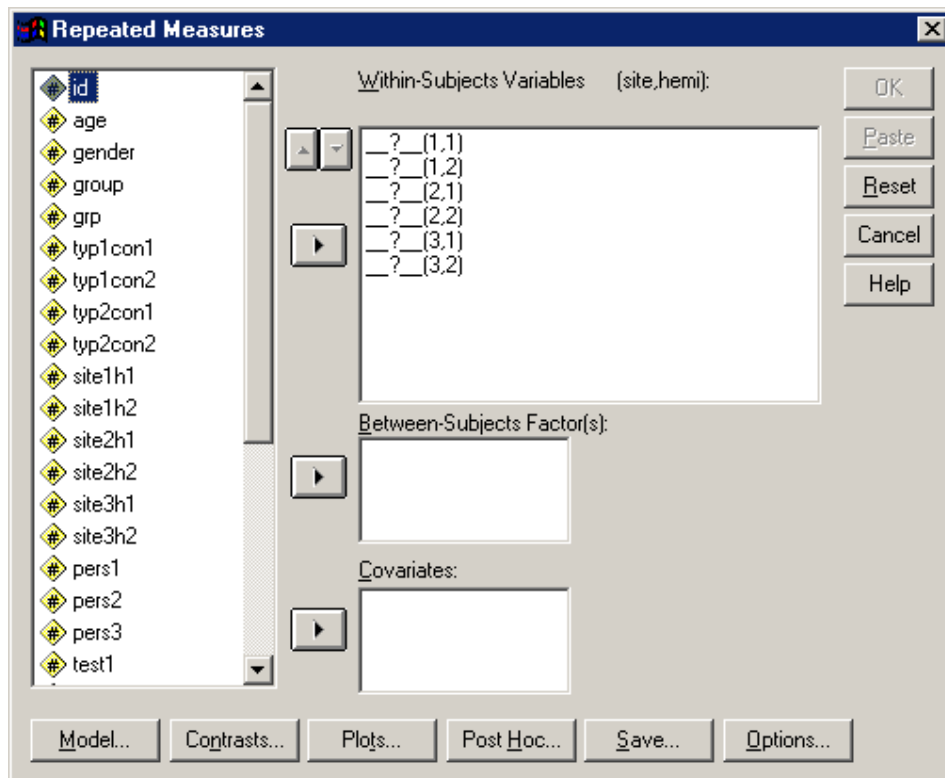
Click on *Analyze* → *General Linear Models* → *Repeated Measures* to get this display:




Enter *site* as the name of the first factor, and specify 3 levels. Click on *Add*. Enter *hemi* as the second factor and specify 2 levels. Click on *Add*. The display should look like this:

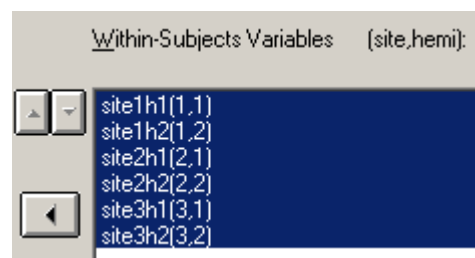


Click on *Define*. The following display will appear:



Notice that the names of the factors are given in order at the top of the display:

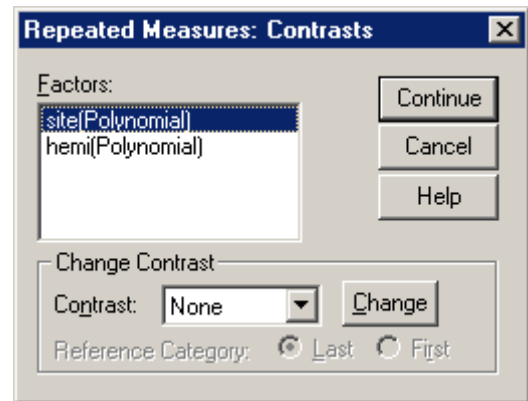
Within-Subjects Variables (site,hemi): . The numbers in brackets in the top window show the levels of each factor. *GLM* expects the first variable to represent *site* level 1 and *hemi* level 1. The next variable is expected to represent *site* level 1 and *hemi* level 2, and so on. Because we specified the factors in the order which corresponds to the order of the variables in the file, we can simply highlight the six variables *site1h1* to *site3h2* from the list at the left of the display and click on the  so that the variables are listed:



(As an exercise, we can ask how the variables would need to be ordered if we had specified *hemi* first, then *site*. This ordering would imply that the three measurements from the first hemisphere come first, with the sites in order 1, 2 and 3, then all three measurements from the second hemisphere. The appropriate order for the variables would be: *site1h1 site2h1 site3h1 site1h2 site2h2 site3h2*.)

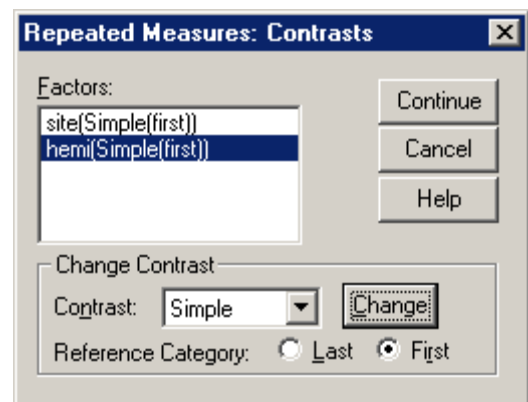
Instead of using the default polynomial contrasts for the two within-subject factors, we will assume that the main questions of interest concern the differences between the first level of each factor and the other levels. We will therefore ask for simple contrasts, with the first category as the reference category. The contrasts will compare the mean for the first level of each factor with the mean for the second level of each factor. And for *site*, the mean for the third level will be compared with the mean for the first level.

To specify these contrasts, click on the *Contrasts* button. This display will appear:



First, click once on *site*, then select *simple* in the *Contrast* slot, and the *First* radio button, then click on the *Change* button. Repeat the process for *hemi*. The display should now look like this:

Click on the *Continue* button.



As always when there is the possibility of an interaction, a graph may be useful. Click on the *Plots* button, and select *site* to be shown on the *Horizontal Axis* and *hemi* to be shown as *Separate Lines*. Click on *Add*.

Finally, click on the *Options* button and check *Estimates of effect size* (this is another good habit to get into) and, for the purposes of this example, check *Transformation matrix*.

#### Syntax

```
glm site1h1 site1h2 site2h1 site2h2 site3h1 site3h2/  
  wsfactor site 3 simple (1) hemi 2 simple (1)/  
  plot=profile(site*hemi)/  
  print=etasq test(mmatrix).
```

- Using syntax, any category can be the reference category, not just the first or last, as with point-and-click.
- The factor given first in the *profile* part of the *plot* subcommand will be represented on the horizontal axis of the graph.

The following are the main parts of the output.

### Within-Subjects Factors

Measure: MEASURE\_1

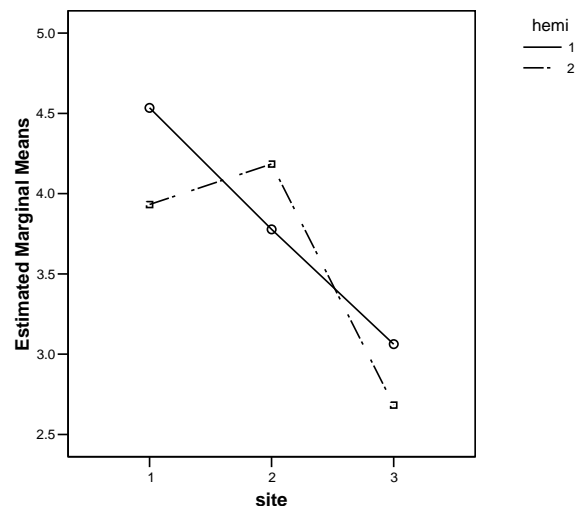
SITE	HEMI	Dependent Variable
1	1	SITE1H1
	2	SITE1H2
2	1	SITE2H1
	2	SITE2H2
3	1	SITE3H1
	2	SITE3H2

The first table (left) shows how the nominated variables fit into the factorial framework. We can see that both the *site1* variables are in the *SIT 1* slot, the *site2* variables are in the *SIT 2* slot, and so on.

Although it is given last in the output, let's look at the graph next (below), to get an idea of what's going on with the data.

Probably the clearest result is that the voltages (the scale is in millivolts) for site 3 tend to be lower than those for the other sites, and this is true for both hemispheres. What's going on at sites 1 and 2 is less clear: perhaps at site 1 the voltages are higher for the left hemisphere, whereas at site 2, the opposite is true. (Or, equivalently, that the difference between sites 1 and 2 is different for the left and right hemisphere.) We're going to need to look at the results of the significance tests to see how seriously to take the patterns shown in the graph.

Estimated Marginal Means of MEASURE\_1



### Box 4: Using pattern rather than colour to differentiate lines in GLM graphs

By default, SPSS uses different-coloured lines and symbols to differentiate treatments or groups in the graphs produced by *GLM*. This is fine on the screen but not much good when you have a monochrome printer. You can always edit the graph before printing it, but there's an easier way: get SPSS to use different line-types and different symbols from the start. To do this, click on *Edit* → *Options* and select the *Charts* tab. At the bottom left of the resulting display, use the pull-down menu and select *Cycle through patterns only*.

The user can also specify the order in which elements such as colours and lines, etc, are used, by clicking on the following buttons:

Style Cycles

Style cycle preference:

Cycle through patterns only  
 Cycle through colors only  
 Cycle through patterns only

Colors...

Lines...

Markers...

Fills...

The next table gives the results of the multivariate analysis. Because *hemi* has two levels, the multivariate and combined univariate results will be exactly the same, but this is not the case for *site*, which has three levels.

**Multivariate Tests<sup>b</sup>**

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
SITE	Pillai's Trace	.537	56.245 <sup>a</sup>	2.000	97.000	.000	.537
	Wilks' Lambda	.463	56.245 <sup>a</sup>	2.000	97.000	.000	.537
	Hotelling's Trace	1.160	56.245 <sup>a</sup>	2.000	97.000	.000	.537
	Roy's Largest Root	1.160	56.245 <sup>a</sup>	2.000	97.000	.000	.537
HEMI	Pillai's Trace	.051	5.269 <sup>a</sup>	1.000	98.000	.024	.051
	Wilks' Lambda	.949	5.269 <sup>a</sup>	1.000	98.000	.024	.051
	Hotelling's Trace	.054	5.269 <sup>a</sup>	1.000	98.000	.024	.051
	Roy's Largest Root	.054	5.269 <sup>a</sup>	1.000	98.000	.024	.051
SITE * HEMI	Pillai's Trace	.209	12.797 <sup>a</sup>	2.000	97.000	.000	.209
	Wilks' Lambda	.791	12.797 <sup>a</sup>	2.000	97.000	.000	.209
	Hotelling's Trace	.264	12.797 <sup>a</sup>	2.000	97.000	.000	.209
	Roy's Largest Root	.264	12.797 <sup>a</sup>	2.000	97.000	.000	.209

a. Exact statistic

b.

Design: Intercept

Within Subjects Design: SITE+HEMI+SITE\*HEMI

The results in the above table show significant effects for *site* and *hemi*, but there is also an interaction between the two factors, so that will be the focus of the interpretation. The partial  $\eta^2$  of .209 shows that the contribution of the interaction is not trivial. The most-often quoted multivariate statistic, Wilks' Lambda, has a value of .791, which shows the proportion of variance not accounted for by *site\*hemi*. The effect size shown in the final column of the table is therefore simply  $1 - .791 = .209$ .

**Mauchly's Test of Sphericity<sup>b</sup>**

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
SITE	.704	34.046	2	.000	.772	.782	.500
HEMI	1.000	.000	0	.	1.000	1.000	1.000
SITE * HEMI	.818	19.436	2	.000	.846	.860	.500

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

a. 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.

b.

Design: Intercept

Within Subjects Design: SITE+HEMI+SITE\*HEMI

In the next table, we once again grapple with the question of whether the assumptions of the univariate approach to repeated measures are met. At least it's only *site* we have to worry about. The significant results of the Mauchly test for *site* and for *site\*hemi*, together with the values of the Greenhouse-Geisser and Huynh-Feldt epsilons, which both clearly depart from 1.00, indicates that we should either use the multivariate results and/or the adjusted univariate results.

## Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
SITE	Sphericity Assumed	207.363	2	103.681	82.995	.000	.459
	Greenhouse-Geisser	207.363	1.543	134.372	82.995	.000	.459
	Huynh-Feldt	207.363	1.563	132.657	82.995	.000	.459
	Lower-bound	207.363	1.000	207.363	82.995	.000	.459
Error(SITE)	Sphericity Assumed	244.852	196	1.249			
	Greenhouse-Geisser	244.852	151.233	1.619			
	Huynh-Feldt	244.852	153.189	1.598			
	Lower-bound	244.852	98.000	2.498			
HEMI	Sphericity Assumed	5.446	1	5.446	5.269	.024	.051
	Greenhouse-Geisser	5.446	1.000	5.446	5.269	.024	.051
	Huynh-Feldt	5.446	1.000	5.446	5.269	.024	.051
	Lower-bound	5.446	1.000	5.446	5.269	.024	.051
Error(HEMI)	Sphericity Assumed	101.286	98	1.034			
	Greenhouse-Geisser	101.286	98.000	1.034			
	Huynh-Feldt	101.286	98.000	1.034			
	Lower-bound	101.286	98.000	1.034			
SITE * HEMI	Sphericity Assumed	27.821	2	13.910	17.834	.000	.154
	Greenhouse-Geisser	27.821	1.693	16.436	17.834	.000	.154
	Huynh-Feldt	27.821	1.719	16.182	17.834	.000	.154
	Lower-bound	27.821	1.000	27.821	17.834	.000	.154
Error(SITE*HEMI)	Sphericity Assumed	152.876	196	.780			
	Greenhouse-Geisser	152.876	165.880	.922			
	Huynh-Feldt	152.876	168.483	.907			
	Lower-bound	152.876	98.000	1.560			

It can be seen from the above table that, although the departure from sphericity for the *site\*hemi* effect is quite marked, this has little practical effect: no matter what adjustment is made, or whether we use the multivariate results instead of the univariate, the interaction is clearly significant.

Before looking at the results for the simple contrasts, we'll check the transformation matrices to see how *GLM* translated our request. We won't consider the matrix which shows how *GLM* created the sum of all six variables for use in the between-subject part of the analysis because this is of no interest in this case.

## Transformation Coefficients (M Matrix)

SITE<sup>a</sup>

Measure: MEASURE\_1

Dependent Variable	SITE	
	Level 2 vs. Level 1	Level 3 vs. Level 1
SITE1H1	-.500	-.500
SITE1H2	-.500	-.500
SITE2H1	.500	.000
SITE2H2	.500	.000
SITE3H1	.000	.500
SITE3H2	.000	.500

a. The contrasts for the within subjects factors are:  
SITE: Simple contrast

HEMI<sup>a</sup>

Measure: MEASURE\_1

Dependent Variable	HEMI
	Level 2 vs. Level 1
SITE1H1	-.333
SITE1H2	.333
SITE2H1	-.333
SITE2H2	.333
SITE3H1	-.333
SITE3H2	.333

a. The contrasts for the within subjects factors are:  
HEMI: Simple contrast

**SITE \* HEMI**

Measure: MEASURE\_1

Dependent Variable	SITE	
	Level 2 vs. Level 1	Level 3 vs. Level 1
	HEMI	HEMI
	Level 2 vs. Level 1	Level 2 vs. Level 1
SITE1H1	1	1
SITE1H2	-1	-1
SITE2H1	-1	0
SITE2H2	1	0
SITE3H1	0	-1
SITE3H2	0	1

a. The contrasts for the within subjects factors are:

SITE: Simple contrast

HEMI: Simple contrast

The coefficients in the upper-left table are normalised (i.e., the sum of the squared coefficients is one in both cases). However, notice that the two simple contrasts are not orthogonal, i.e.,  $(-.5 * -.5) + (-.5 * -.5) + \dots + (0 * .5) = .5$ , rather than 0.00. Unlike the *manova* procedure, *GLM* doesn't insist that contrasts of within-subject factors are orthogonal. Finally, the coefficients in the table at left, for the interaction, are not only not orthogonal, they are not normalised.

The signs of the coefficients in the table at left are simply a result of multiplying the coefficients for *site* by the corresponding coefficients for *hemi*.

However, we can also work out what

they should be by considering the contrasts we requested. The first *site* contrast compares *site2* with *site1*. The convention is that the contrast coefficient for the reference category or level is negative. A negative difference therefore means that the mean for the reference category is higher than that for the category which is being compared to it. The *hemi* contrast compares *hemi 1* with *hemi 2*. The interaction contrast can be thought of in either of two ways: (1) Is  $[site2h2 - site2h1] - [site1h2 - site1h1] = 0$ , i.e., is the difference between the left and right hemisphere different for site 1 and site2? and (2) Is  $[site2h2 - site1h2] - [site2h1 - site1h1] = 0$ , i.e., is the difference between site 1 and site 2 different for the two hemispheres? These two equations can be given in the following form:

$$(1) \text{ site2h2} - \text{site2h1} - \text{site1h2} + \text{site1h1}$$

$$(2) \text{ site2h2} - \text{site1h2} - \text{site2h1} + \text{site1h1}.$$

The signs are therefore +1, -1, -1, +1, which are the signs of the coefficients shown in the first column of the last table above.

The table of within-subject contrasts, shown next, reflects the request for simple comparisons of each mean (e.g., *site2* and *site3* for *site* and *hemi2* for *hemi*) with the mean for the first level (*site 1* for *site* and the left hemisphere for *hemi*).

Tests of Within-Subjects Contrasts

Measure: MEASURE\_1

Source	SITE	HEMI	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
SITE	Level 2 vs. Level 1		6.303	1	6.303	8.842	.004	.083
	Level 3 vs. Level 1		183.201	1	183.201	97.502	.000	.499
Error(SITE)	Level 2 vs. Level 1		69.860	98	.713			
	Level 3 vs. Level 1		184.137	98	1.879			
HEMI		Level 2 vs. Level 1	3.631	1	3.631	5.269	.024	.051
Error(HEMI)		Level 2 vs. Level 1	67.524	98	.689			
SITE * HEMI	Level 2 vs. Level 1	Level 2 vs. Level 1	100.792	1	100.792	25.729	.000	.208
	Level 3 vs. Level 1	Level 2 vs. Level 1	4.905	1	4.905	2.725	.102	.027
Error(SITE*HEMI)	Level 2 vs. Level 1	Level 2 vs. Level 1	383.910	98	3.917			
	Level 3 vs. Level 1	Level 2 vs. Level 1	176.399	98	1.800			

The first interaction contrast in the above table is clearly significant, whatever adjustment we might care to make: the difference between the two hemispheres is different for sites 1 and 2, and the difference between sites 1 and 2 is different for the left and right hemispheres, whichever way we'd like to put it. However, as would be expected from the graph, the second contrast is not significant, i.e., the difference between the left and right hemisphere isn't different for sites 1 and 3 (or equivalently ....?).

Note: The contrasts discussed above could also be tested with the following *mmatrix* syntax:

```
mmatrix="s2vs1 for h2 vs h1" site1h1 -1 site1h2 1 site2h1 1 site2h2 -1 site3h1 0 site3h2 0;
"s3vs1 for h2 vs h1" site1h1 -1 site1h2 1 site2h1 0 site2h2 0 site3h1 1 site3h2 -1.
```

However, for point-and-click users, the method described earlier is clearly preferable. And p-and-clickers will need to save their resources for designs which include both between-and within-subject factors, because if the between-subject factor has more than two levels, the only way of obtaining interaction contrasts with them is to use *lmatrix* and *mmatrix* together.

### 7.1 Tests of Simple Effects (syntax only)

The other way of following up a significant interaction is to test the simple effects, which in this case means testing the effect of *site* separately for each hemisphere and testing the effect of *hemi* separately for each site. This is easily done using the *emmeans* subcommand:

```
emmeans=table(site*hemi) compare(site)/
emmeans=table(site*hemi) compare(hemi).
```

These commands produce the following tables:



## Pairwise Comparisons

Measure: MEASURE\_1

HEMI	(I) SITE	(J) SITE	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
						Lower Bound	Upper Bound
1	1	2	.757*	.169	.000	.422	1.092
		3	1.472*	.155	.000	1.165	1.779
	2	1	-.757*	.169	.000	-1.092	-.422
		3	.715*	.130	.000	.458	.972
	3	1	-1.472*	.155	.000	-1.779	-1.165
		2	-.715*	.130	.000	-.972	-.458
2	1	2	-.252*	.075	.001	-.401	-.103
		3	1.249*	.152	.000	.947	1.551
	2	1	.252*	.075	.001	.103	.401
		3	1.501*	.158	.000	1.188	1.815
	3	1	-1.249*	.152	.000	-1.551	-.947
		2	-1.501*	.158	.000	-1.815	-1.188

Based on estimated marginal means

\*. The mean difference is significant at the .050 level.

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

At first glance, every mean seems to be different from every other mean. At second glance, they still are. If we treat these comparisons as *post hoc* ones, and adjust for the six comparisons involved ( $3!/((3-2)!2!) = 3$ , times two hemispheres) to keep the family-wise error rate at .05, the critical value is  $.05/6 = .008$ . All of the *p*-values in the above table are therefore significant.

## Pairwise Comparisons

Measure: MEASURE\_1

SITE	(I) HEMI	(J) HEMI	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
						Lower Bound	Upper Bound
1	1	2	.602*	.108	.000	.389	.815
	2	1	-.602*	.108	.000	-.815	-.389
2	1	2	-.407*	.179	.025	-.762	-.5.207E-02
	2	1	.407*	.179	.025	5.207E-02	.762
3	1	2	.379*	.094	.000	.193	.566
	2	1	-.379*	.094	.000	-.566	-.193

Based on estimated marginal means

\*. The mean difference is significant at the .050 level.

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

The adjusted *p*-value for the comparisons of hemispheres (one at each of the three sites) is  $.05/3 = .0167$ . By this criterion, the only difference which isn't significant is the one at site 2. This isn't something which is obvious from the graph.



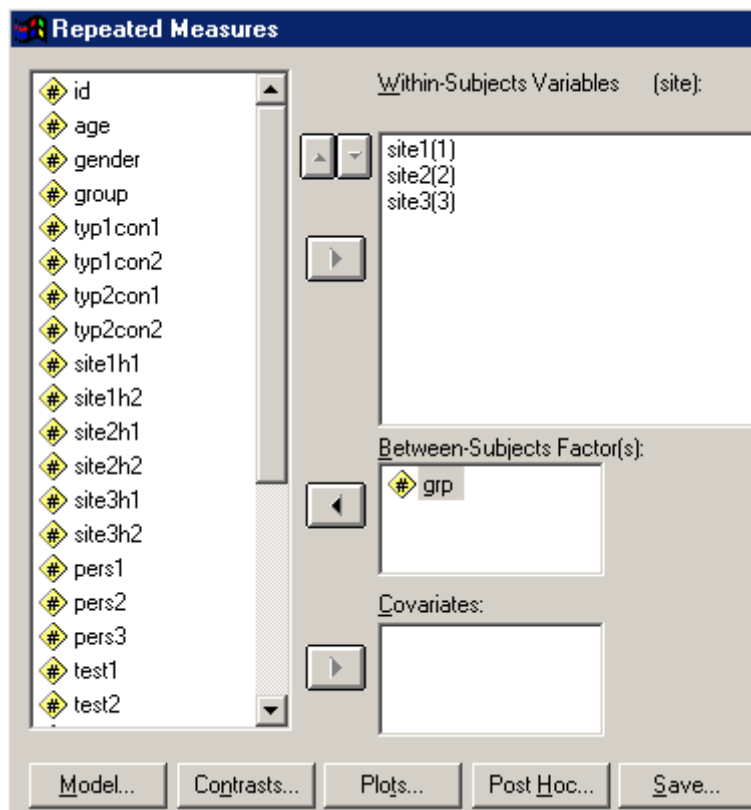
## 8. A Mixed Design: One Within- and One Between-Subject Factor

Perhaps the most common kind of ANOVAs in Psychology are those which contain both between- and within-subject factors. These are 'mixed' in two senses – they contain both between- and within-subject factors, as mentioned, and they contain both fixed and random factors. In this section, we'll be directly concerned only with the fixed factors – in the analyses covered here, *GLM* looks after the random factor of subject for us. However, it is worth noting that in the analyses referred to in Section 10 (and in other analyses you might perform in SPSS), the subject factor becomes explicit.

The example we'll look at contains one between-subject factor, *grp*, with two levels (control and experimental), and one within-subject factor, *site*, with three levels. The three site variables, *site1*, *site2* and *site3*, were created from the variables used in the previous analysis by averaging over hemisphere:

```
compute site1=mean(site1h1,site1h2).
compute site2=mean(site2h1,site2h2).
compute site3=mean(site3h1,site3h2).
```

Specify the basic analysis by clicking *Analyze* → *General Linear Model* → *Repeated Measures*. Specify *site* as the within-subject factor, with three levels. On the main *Repeated Measures* display, enter *site1*, *site2* and *site3* as the *Within-Subjects Variables*., and *grp* as the *Between-Subjects Factor*. The display should look like this:



Since the analysis will automatically test the interaction between *site* and *grp*, request a plot, with *site* on the horizontal axis and a separate line for the two groups (*grp*).

## Syntax

```
glm site1 site2 site3 by grp/
  wsfactor site 3/
  plot=profile(site*grp).
```

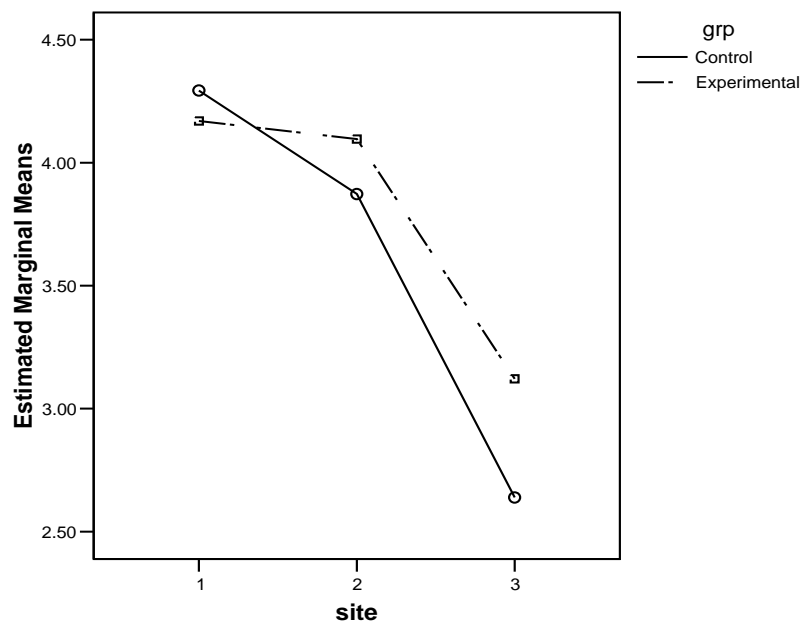
For our purposes, we'll look only at the univariate results and the graph. (It's worth noting, though, that this is one case where the univariate and multivariate results disagree: for  $site*grp$ , Wilks' Lambda = .943,  $F(2, 96) = 2.91$ ,  $p = .059$ , whereas, as can be seen in the table below, all the univariate results except the worst-case lower-bound give  $p < .05$ .)

## Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SITE	Sphericity Assumed	102.392	2	51.196	84.279	.000
	Greenhouse-Geisser	102.392	1.558	65.716	84.279	.000
	Huynh-Feldt	102.392	1.595	64.187	84.279	.000
	Lower-bound	102.392	1.000	102.392	84.279	.000
SITE * GRP	Sphericity Assumed	4.580	2	2.290	3.770	.025
	Greenhouse-Geisser	4.580	1.558	2.939	3.770	.035
	Huynh-Feldt	4.580	1.595	2.871	3.770	.034
	Lower-bound	4.580	1.000	4.580	3.770	.055
Error(SITE)	Sphericity Assumed	117.846	194	.607		
	Greenhouse-Geisser	117.846	151.135	.780		
	Huynh-Feldt	117.846	154.737	.762		
	Lower-bound	117.846	97.000	1.215		

## Estimated Marginal Means of MEASURE\_1



### 8.1 Interaction Contrasts in a Mixed Analysis using *contrast*

For this example, we'll assume that, given there is a significant interaction, there are two interaction contrasts that interest us: (1) Whether the difference between the control and experimental groups is different for sites 1 and 2, and (2) Whether the difference between site 3 and the average of sites 1 and 2 is different for the control and experimental groups. As it happens, the contrasts of *site* that we're interested in are *difference* contrasts, in which each level is compared to the average of the preceding levels. In this example, then, we can click on the *Contrasts* button and ask for *Difference* contrasts for *site*. I say 'in this example' because the method we are about to use works only if the between-subject factor has only two levels; if the between-subjects factor has three or more levels, the individual one-degree-of-freedom contrasts are lost in the output given by *GLM*, because it combines the results for all of the between-subjects contrasts (which would be three with the four-level *group* variable, for example).

Set up the analysis as before, but also click on *Contrasts*, and specify *Difference* for *site* and *Simple/First* for *grp* (although in this case the results will be the same whatever contrast is chosen for *grp*).

#### Syntax

```
glm site1 site2 site3 by grp/  
wsfactor site 3 difference/  
contrast(grp)=simple(1).
```

#### Tests of Within-Subjects Contrasts

Measure: MEASURE\_1

Source	site	Type III Sum of Squares	df	Mean Square	F	Sig.
site	Level 2 vs. Level 1	6.038	1	6.038	8.757	.004
	Level 3 vs. Previous	149.060	1	149.060	114.196	.000
site * grp	Level 2 vs. Level 1	2.987	1	2.987	4.333	.040
	Level 3 vs. Previous	4.630	1	4.630	3.547	.063
Error(site)	Level 2 vs. Level 1	66.873	97	.689		
	Level 3 vs. Previous	126.614	97	1.305		

The table *Tests of Within-Subjects Contrasts* shows that the *p*-values for the test of whether the difference between Site 1 and Site 2 is different for *grp* = 1 and *grp* = 2 is .040, and for the test of whether the difference between the average of Site 1 and Site 2, and Site 3, is different for the two groups is .063. The calculation of the contrasts is described in the next section, where another method of testing interaction contrasts in the mixed model is given; one which applies when there are more than two between-subject groups. In this connection, a final point is that if our analysis had involved the four-level *group* rather than the two-level *grp*, the above table would have looked the same (except that the sums of squares and *p*-values would have been different) because *GLM* would not have distinguished between the three *group* contrasts and would have simply lumped them together for the purposes of the *Level 2 vs. Level 1* part (for example) of the *site \* grp* part of the table.

## 8.2 Interaction Contrasts in a Mixed Analysis using *lmatrix* and *mmatrix* (syntax only)

Because there are only two between-subject groups, and we're interested in comparing them to each other, rather than (for example) to the overall mean, a simple contrast with the lowest-numbered group (the controls) as the reference group is appropriate. The *lmatrix* subcommand is then:

```
lmatrix "Exp v Cont" grp -1 1/
```

Although *GLM* allows us to specify more, we'll limit ourselves to two within-subject contrasts. The *mmatrix* subcommand is:

```
mmatrix "Site2 v Site 1" site1 -1 site2 1 site 3 0; "Site3 v Site12" site1 -.5 site2 -.5 site3 1
```

Notice that all we've done is specify the appropriate contrast(s) for each factor: we haven't done anything about the actual interaction contrasts, which we would have to do if both factors were within-subject factors or both factors were between-subject factors. The good news is that *GLM* does this for us – as with the factors themselves, *GLM* automatically tests interactions between within- and between-subject factors.

### Syntax

```
glm site1 site2 site3 by grp/  
wsfactor site 3/  
print=descriptives/  
lmatrix "grp2 v 1" grp -1 1/  
mmatrix "site2 v 1" site1 -1 site2 1 site3 0; "site3 v12" site1 -.5 site2 -.5 site3 1.
```

The *print=descriptives* subcommand has been added so we can check the results of the interaction contrasts.

Note: For future reference, if the analysis involved the four-level factor of *group*, the *lmatrix* command

```
lmatrix "group pairwise contrasts" group -1 1 0 0; group -1 0 1 0; group -1 0 0 1;  
group 0 -1 1 0; group 0 -1 0 1; group 0 0 -1 1/
```

would carry out the interaction contrasts involving all pairwise comparisons of *group*.

Because *GLM* always tests contrasts involving within-subject factors against error terms based on just the variables involved in the contrast, not on a pooled error term (as used in the univariate approach to repeated measures), we're not going to worry about the tests of sphericity, so the only relevant tables in the output are:

**Descriptive Statistics**

	GRP	Mean	Std. Deviation	N
SITE1	1 Control	4.2935	.81765	51
	2 Experimental	4.1695	.79070	48
	Total	4.2334	.80301	99
SITE2	1 Control	3.8727	.58168	51
	2 Experimental	4.0962	.75795	48
	Total	3.9810	.67878	99
SITE3	1 Control	2.6391	.77240	51
	2 Experimental	3.1216	1.08596	48
	Total	2.8730	.96369	99

**Custom Hypothesis Tests Index**

1	Contrast Coefficients (L' Matrix)	LMATRIX Subcom mand 1: grp2 v 1
	Transformation Coefficients (M Matrix)	MMATRIX Subcom mand
	Contrast Results (K Matrix)	Zero Matrix

**Custom Hypothesis Tests #1****Contrast Results (K Matrix)<sup>a</sup>**

Contrast		Transformed Variable	
		site2 v 1	site3 v12
L1	Contrast Estimate	.3476	.4327
	Hypothesized Value	.0000	.0000
	Difference (Estimate - Hypothesized)	.3476	.4327
	Std. Error	.1670	.2298
	Sig.	.0400	.0627
	95% Confidence Interval Lower Bound	.0162	-.0233
	for Difference Upper Bound	.6790	.8887

a. Based on the user-specified contrast coefficients (L') matrix: grp2 v 1

The first table gives the means for each combination of factors. The second table is an index of the 'custom' contrasts we've requested. Although this index is a little obscure, it at least tells the reader that the first set of contrasts involved both the *lmatrix* and the *mmatrix*; so it's clear that we're dealing with interaction contrasts between a within- and between-subject factor.

The first question we had was whether the difference between the control and experimental groups is different for sites 1 and 2. The question is whether  $(\text{site2 for grp 2} - \text{site2 for grp 1}) - (\text{site1 for grp 2} - \text{site1 for grp 1}) = 0$ . Using the means from the above table, the calculation is  $(4.0962 - 3.8727) - (4.1695 - 4.2935) = .3475$ , which is, within rounding error, what is shown in the above table, which is also shown to have a *p*-value of .04 (the same value as that given in the previous section). If we were using a decision-wise criterion of .05, this would be a statistically significant result. However,

we'll assume that these two interaction contrasts are *a priori* and that we want to maintain the Type I error rate at .05 over both contrasts, so that the critical value is .025, making the contrast non-significant.

The second question was whether the difference between site 3 and the average of sites 1 and 2 is different for the control and experimental groups. The calculation is  $(3.1216 - .5 * 4.1695 - .5 * 4.0962) - (2.6391 - .5 * 4.2935 - .5 * 3.8727) = .4328$ . This difference is also not significant, at  $p = .0627$  (again, the same value was obtained in the previous section).

### 8.3 Test of Simple Effects in a Mixed Analysis (syntax only)

Testing simple effects after an interaction is as simple in a mixed design as it is when both factors are the same kind. The subcommands to test all simple effects in the present example are:

```
emmeans=table(grp*site) compare(grp)/
emmeans=table(grp*site) compare(site).
```



## 9. A Numeric Independent Variable in a Repeated Measures Design

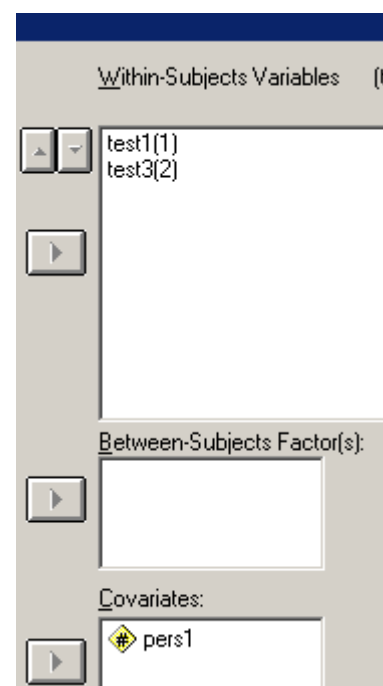
Traditionally, the between-subject independent variables in repeated measures analyses are categorical factors rather than numeric variables. Therefore, users contemplating such analyses often feel an overwhelming urge to categorise numeric variables in order to conform to this model. They do this even when they acknowledge that there is really no good reason to waste the information in the numeric variable by lumping together cases with different values which may reflect real differences between subjects. *GLM* makes it possible to include numeric variables in repeated measures analyses. This section first considers the use of numeric variables as independent variables and then their use as covariates when the real interest is in between- or within-subject factors.

### 9.1 Interpreting the Effect of a Numeric Independent Variable in a Repeated Measures Analysis

In a mixed design ANOVA, with a between-subject factor and a within-subject factor, the question of most interest is usually whether there is an interaction between the two factors. For example, when the within-subject factor is time, and the between-subject factor is treatment (e.g., experimental versus control), the question is whether the change over time is different for the treatments. With two treatment groups, the test of the interaction is equivalent to an independent groups *t*-test with the difference score (e.g., time 2 – time 1) as the dependent variable. If, instead of having group as the independent variable, we have a numeric variable, the question then is whether there is a linear relationship between the difference score and the independent variable. For example, it may be that among subjects whose self-esteem is measured at baseline and then measured again after they all attend a course called 'Appreciating the Self – A Cognitive Approach', older subjects' self-esteem increases more than that of younger subjects. In this case, we would expect (if a high score on the test equals high self-esteem, and the time 1 score is subtracted from the time 2 score) a positive correlation between age and the difference score.

As a concrete example, we'll use two *test* variables, *test1* and *test3*, with *pers1* as the numeric independent variable. The question is whether the change from *test1* to *test3* is related to the personality variable *pers1*. For point-and-click, click on *Analyze* → *General Linear Model* → *Repeated Measures*, then specify *test*, with two levels, as the within-subject factor, click on *Define*, and specify *test1* as the first within-subject variable and *test3* as the second, and *pers1* as a covariate. The display should look like this:

For future use, click on *Options*, then check *Estimates of effect size*.



## Syntax

```
glm test1 test3 with pers1/
wsfactor test 2/
print=etasq.
```

For our purposes, we'll use the univariate output (because there are only two levels in the

## Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
TEST	Sphericity Assumed	4.219	1	4.219	7.406	.008	.071
	Greenhouse-Geisser	4.219	1.000	4.219	7.406	.008	.071
	Huynh-Feldt	4.219	1.000	4.219	7.406	.008	.071
	Lower-bound	4.219	1.000	4.219	7.406	.008	.071
TEST * PERS1	Sphericity Assumed	3.036	1	3.036	5.330	.023	.052
	Greenhouse-Geisser	3.036	1.000	3.036	5.330	.023	.052
	Huynh-Feldt	3.036	1.000	3.036	5.330	.023	.052
	Lower-bound	3.036	1.000	3.036	5.330	.023	.052
Error(TEST)	Sphericity Assumed	55.257	97	.570			
	Greenhouse-Geisser	55.257	97.000	.570			
	Huynh-Feldt	55.257	97.000	.570			
	Lower-bound	55.257	97.000	.570			

within-subject factor, the multivariate results are exactly the same). The significant result for *test* in the above table is not of immediate interest: it shows that there is a significant difference between *test1* and *test3* (adjusted for *pers1*, but we'll talk about that in the next section). The main interest is in the interaction, *test\*pers1*, which is significant ( $F(1,97) = 5.33, p = .023$ ). This means that there is a systematic linear relationship between *pers1* and the difference between *test1* and *test3*. What's the nature of this relationship? It's no use asking *GLM* for parameter estimates, because all we get are the regression coefficients for the relationship between *pers1* and each of *test1* and *test3*, and nothing for the relationship between *pers1* and the difference between *test1* and *test3*. One thing to do to is to work out the correlation between *pers1* and the difference. The following syntax does this:

```
compute t31diff=test3 - test1.
correlations pers1 t31diff.
glm t31diff with pers1/
print=parameters.
```

## Correlations

		PERS1 Personality score 1	T31DIFF
PERS1 Personality score 1	Pearson Correlation	1	.228
	Sig. (2-tailed)	.	.023
	N	99	99
T31DIFF	Pearson Correlation	.228	1
	Sig. (2-tailed)	.023	.
	N	99	99

The correlation is .228. Because the difference score is *test3* – *test1*, this means that higher values of *pers1* are associated with higher values of *test3* relative to *test1*: the value of *test* increases more (or decreases less) for subjects with higher values of

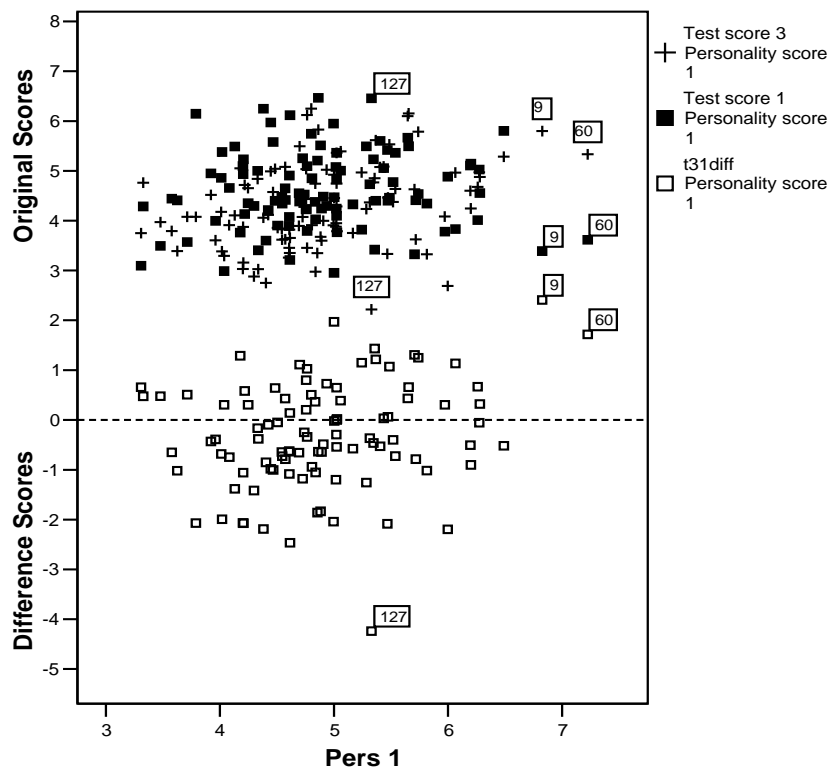
*pers1*. Notice that, as would be expected, the  $p$ -value for the correlation, .023, is exactly the same as that for the *test* by *pers1* interaction in the ANOVA and, again as expected, the correlation of .228 is the square root of the  $\eta^2$  given in the ANOVA table.

#### Parameter Estimates

Dependent Variable: T31DIFF

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	-1.907	.701	-2.721	.008	-3.297	-.516
PERS1	.326	.141	2.309	.023	4.578E-02	.607

Calculating the correlation between *pers1* and the difference score is an exactly equivalent analysis to the ANOVA. The second supplementary analysis, the regression of the difference score on *pers1*, is also exactly equivalent. As can be seen in the above table, the  $p$ -value is again .023, but the table also tells us that for every unit increase in *pers1*, the difference score increases by .326. Despite all this information, I don't feel I've got a real grip on the situation unless I can see it; so a graph is called for:



This rather complicated plot shows the original *test1* (filled squares) and *test3* (+) in the top part of the graph and the difference scores (open circles) in the bottom part. The positive relationship between *pers1* and the difference scores can be seen but, embarrassingly, it can also be seen that it may be due mainly to two extreme cases, labelled '9' and '60' in the graph. Worst fears are confirmed when the analysis is repeated without these two subjects: the *test*\**pers1* interaction is not now significant ( $r = .12$ ,  $p = .243$ ). Omitting the other extreme case, '127', makes no further difference. Hmm. In an actual analysis we would have to decide at this point whether we want to report a result which depends on a few extreme cases or report that there is no strong support for the hypothesis that change from *test1* to *test3* is related to *pers1*. Because this is an

illustration only, we're not concerned to sort out this issue here. We can be thankful that, as well as allowing us to see how a quantitative independent variable can be used in a repeated measures analysis, these data have shown the value of graphing results (and in general, poking around in the data) in order to pick up issues which may not become obvious in any other way.

## 9.2 Analysis of Covariance in Repeated Measures Designs – Adjusting the Effect of a Between-Subjects Factor

In the previous section, the relationship between a numeric variable, *pers1*, and the difference between *test1* and *test3* was assessed. This leads naturally onto using a numeric variable as a covariate in a mixed analysis. As an example we'll continue with the same within-subject factor (*test*) and the same numeric variable (*pers1*) but add the two-level between-subject factor, *grp*. The question is whether any interaction between *grp* and *test* can be accounted for by the interaction between *pers1* and *test*. For this to be a possibility, there has to be a relationship between *pers1* and *test* (which, from the previous section, we know there is, however spurious) and the groups must have different means on *pers1*. And, for us to be interested, there will probably be a significant interaction between *grp* and *test* to start with. Some of these conditions are met only weakly with the present data, but they will suffice for illustration. The table below shows that the experimental group has a higher mean than the control group on *pers1* (4.98 versus 4.83) and also, that the extent of negative change is smaller for the experimental group (-.1169) than for the control group (-.4883). The proposition that we'll test

Group Statistics

GRP		N	Mean	Std. Deviation	Std. Error Mean
PERS1 Personality score 1	1 Control	51	4.83	.661	.093
	2 Experimental	48	4.98	.858	.124
T31DIFF	1 Control	51	-.4883	1.12703	.15782
	2 Experimental	48	-.1169	1.02797	.14837

is that any interaction between *grp* and *test*, which suggests that the subjects in the experimental group experienced a smaller negative change between *test1* and *test3* than those in the control group, occurred because they were higher on *pers1*, larger values of which are associated with less negative change.

First, is there an interaction between *grp* and *test*? The results of the mixed-model analysis are shown in the following table (since circularity isn't an issue with only two levels in the within-subjects factor, only the sphericity-assumed results are shown). For the point-and-click commands see Section 8.

### Syntax

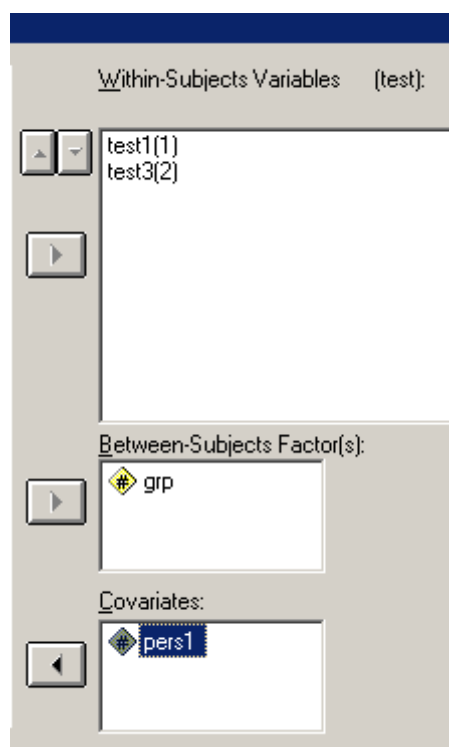
```
glm test1 test3 by grp/
wsfactor test 2.
```

**Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	4.529	1	4.529	7.763	.006
TEST * GRP	1.706	1	1.706	2.924	.090
Error(TEST)	56.588	97	.583		

To justify further analyses, we'll declare that there is a significant interaction between *grp* and *test* at  $p < .10$ . Now the analysis is rerun with *pers1* as the covariate. For point-and-click, the commands are a combination of those in Section 8 and Section 9.1. The display should look like this:

**Syntax**

```
glm test1 test3 by grp with pers1/
wsfactor test 2.
```

The resulting table is:

**Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	3.716	1	3.716	6.611	.012
TEST * PERS1	2.632	1	2.632	4.683	.033
TEST * GRP	1.301	1	1.301	2.315	.131
Error(TEST)	53.956	96	.562		

The  $p$ -value for the  $test*grp$  interaction is now not significant at the .10 level, which supports the suggestion that the apparent effect of  $grp$  on the difference between  $test1$  and  $test2$  may be due to differences between the groups in terms of  $test3$ .

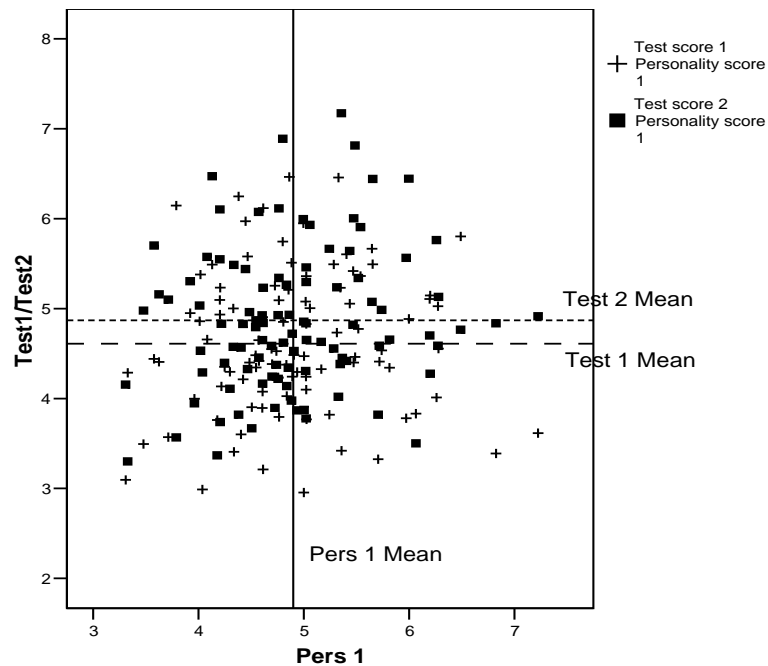
### 9.3 Analysis of Covariance in Repeated Measures Designs – Adjusting the Effect of a Within-Subjects Factor

There are two ways of looking at an analysis of covariance with a single within-subjects factor (as opposed to the analysis of covariance and between-subject factors in a mixed model analysis, as discussed in the previous section). One is to say that if (for example) there are three levels in the within-subject factor, a covariate isn't relevant unless it is also measured on three occasions, each measurement being made at the same time as that for the corresponding level of the within-subject factor. This kind of covariate is known as a **time-varying covariate**, even though the levels of the within-subject factor may not be different times. In a psychophysical experiment, for example, subjects may be measured on three different tasks, with unavoidable changes in the level of illumination for each task. If illumination levels were measured at the same time as the performance at each of the three times, they could be used as covariates. The differences between tasks could be assessed while statistically holding illumination constant. In this approach, knowing the illumination for the first task would not be helpful when it came to comparing performance on different tasks.

The second approach to analysis of covariance with within-subject factors, on the other hand, recognises what was discussed in Section 9.1 – that the magnitude of change over levels of a within-subjects factor may be related to a single variable. This kind of variable is known as a **subject-based covariate**, which is measured on only one occasion and remains constant for that subject. In the example in the previous section, the focus was on the relationship between the independent variable ( $pers1$ ) and the difference between the  $test3$  and the  $test1$  scores -- the  $test*pers1$  interaction (never mind that the relationship may have been spurious). In that analysis no attention was paid to the  $test$  main effect, and how the inclusion of  $pers1$  could have affected its magnitude. We'll now consider an example similar to the previous one but in which the interest is in the  $test$  effect and whether it's altered by a covariate. Following that, we'll return to deal with the first way of looking at analysis of covariance in repeated measures analyses.

#### 9.3.1 A Within-Subject Factor and a Single (Subject-Based) Covariate

This example involves a within-subject factor with two levels, represented by  $test1$  and  $test2$ . The covariate is  $pers1$ , used in the previous example. Back in Section 6.1, a paired  $t$ -test and repeated measures analysis were carried out with  $test1$  and  $test2$ . These analyses showed that there was a significant difference between the means of the two variables,  $p = .021$ . Building on the experience of Section 9.1, and taking nothing for granted, we'll look at a graph of this result which will also come in handy when getting to grips with the effect of the covariate.



Ignoring the covariate in the meantime, the points for *test1* and *test2*, and their respective means, shown by the horizontal dashed (*test1*) and dotted (*test2*) lines, definitely suggest that there is a difference between the two variables (inspection of the graph shows that there are more *test1* points 'sticking out' at the bottom of the point cloud and more *test2* points 'sticking out' at the top of the cloud).

Another way of looking at this analysis, one which will be useful later, is to say that we are simply testing whether the difference between *test1* and *test2* is different from zero. The analysis can therefore be carried out as follows:

```
compute t21diff=test2 - test1.
glm t21diff/
print=parameters.
```

In other words, we create our own difference score and present *GLM* with it as the only variable in a regression analysis. Instead of crashing or asking us what we think we're doing, *GLM* produces this output:

#### Tests of Between-Subjects Effects

Dependent Variable: T21DIFF

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.000 <sup>a</sup>	0	.	.	.
Intercept	6.372	1	6.372	5.500	.021
Error	113.544	98	1.159		
Total	119.917	99			
Corrected Total	113.544	98			

a. R Squared = .000 (Adjusted R Squared = .000)

*GLM* has fitted a model with only a constant,  $t21diff = B_0$ , and performed a test of whether the mean difference is significantly different from zero. The result of the test is

exactly the same as that for the 'official' repeated measures test ( $p = .021$ ). This is essentially what the test of a main effect of a within-subject factor is about – seeing whether the difference (or other contrast[s]) are different from zero. The parameter estimate is as follows:

**Parameter Estimates**

Dependent Variable: T21DIFF

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	.254	.108	2.345	.021	.039	.468

Notice that the estimate of .254 is equal to the mean difference  $test2 - test1$ .

So, what happens when the covariate *pers1* is included in the conventional repeated measures analysis? Click on *Analyze* → *General Linear Model* → *Repeated Measures*. Specify a within-subjects factor called *test* with two levels. Enter *test1* and *test2* as the within-subjects variables and *pers1* as a covariate.

Syntax

```
glm test1 test2 with pers1/
wsfactor test 2.
```

The output is as follows (because sphericity isn't an issue with two levels in the within-subject factor, only the 'Sphericity Assumed' results are shown):

**Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	.071	1.000	.071	.121	.729
TEST * PERS1	.297	1.000	.297	.511	.477
Error(TEST)	56.475	97.000	.582		

The main thing is that although *pers1* doesn't seem to be related to the difference between *test1* and *test2* ( $p = .477$ ), the difference is no longer significant ( $p = .729$ ). What does this mean? The answer becomes clear when we consider another regression analysis with the difference score as the dependent variable:

```
glm t21diff with pers1/
print=parameters.
```

This is testing the model  $t21diff = B_0 + B_1 * pers1$ . In any equation of this sort, what is the interpretation of  $B_0$ ? It is the predicted value of the dependent variable (the difference score in this case) when *pers1* is zero. The results of this analysis are shown below:



### Tests of Between-Subjects Effects

Dependent Variable: T21DIFF

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.595 <sup>a</sup>	1	.595	.511	.477
Intercept	.141	1	.141	.121	.729
PERS1	.595	1	.595	.511	.477
Error	112.950	97	1.164		
Total	119.917	99			
Corrected Total	113.544	98			

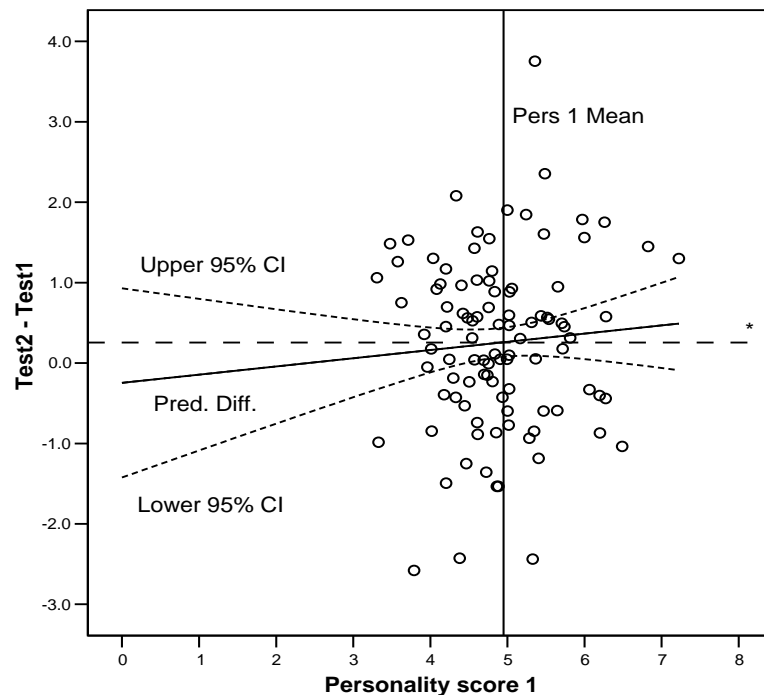
a. R Squared = .005 (Adjusted R Squared = -.005)

### Parameter Estimates

Dependent Variable: T21DIFF

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	-.247	.708	-.348	.729	-1.652	1.159
PERS1	.102	.143	.715	.477	-.181	.386

The first thing to notice is that the  $p$ -values for the Intercept and *pers1* are exactly the same as those for *test* and *test\*pers1* in the repeated measures analysis. We are carrying out the same analysis in a different way. The second thing to observe is that the value of  $B_0$ , shown in the second table as  $-.247$ , must be the predicted value of *t21diff* when the value of *pers1* is zero. A third point to notice is that the standard error for the Intercept in the analysis with *pers1* (.708) is much larger than that for the intercept in the analysis which did not include *pers1* (.108). The reason for this can be shown with the following graph:



The points (open circles) in the graph are the difference scores (*test2* – *test1*). The unadjusted mean of these scores (.254) is shown by the coarsely-dashed line (marked at

its right-hand end by an asterisk). As the previous analyses have shown, this mean is significantly greater than zero. The sloping unbroken line between the curved dashed lines is the regression line showing the relationship between the difference scores and *pers1* and, as can be seen by following the line to the point where *pers1* is equal to zero, the predicted value of the difference scores is around  $-.247$ , the value obtained in the analysis above. The interesting thing is that the difference of  $-.247$  is almost as far below zero as  $.254$  is above it, yet it isn't remotely significant. The reason for this is shown by the curved dashed lines above and below the regression line. These show the error involved in predictions made with the regression line. It is very clear that there is much more error in predictions of the difference score for values of *pers1* which are far from the mean of *pers1* (4.9) than there is for values which are close to the mean (the two confidence lines are closest at the mean of *pers1*).

To summarise: when a single covariate is included in a repeated measures analysis, the within-subject main effect tested is that which would be expected when the covariate has a value of zero. This will often be at best a not-very-useful result, and at worst nonsensical, as in the present case. Further, if zero is quite a long way from the mean of the covariate, the error for the test of whether the within-subject main effect is zero will be very large.

Given the above, it won't normally be sensible to include a single covariate in a repeated measures analysis if the interest is in how it affects the within-subject main effect (it is still perfectly sensible to ask how the covariate is related to the difference between within-subject variables, as discussed in Section 9.1).

One circumstance under which an analysis like the above would be useful is when we would like to see whether the effect of a within-subject factor is significant at a particular value of the covariate. We can use the fact that the within-subject factor is tested at the point where the covariate is zero to achieve this, by using a method called *centring*. Given that scores on *pers1* range from 3 to 7, it would be sensible to ask whether *test* is significant if *pers1* is equal to 4. The following commands would do the trick:

```
temporary.  
compute pers1 = pers1 - 4.  
glm test1 test2 with pers1/  
wsfactor test 2.
```

When 4 is subtracted from *pers1*, there is a new zero point, corresponding to 4 and, because the effect of *test* is assessed at the point where *pers1* equals zero, we have 'fooled' *GLM* into testing the effect of *test* for *pers1* = 4. As an exercise, you might like to verify that if 4.9 (the mean of *pers1*) is subtracted from *pers1*, the result of the above analysis for *test* is identical to that obtained when *pers1* is not included as a covariate.

We will have more to say about centring as a useful tool for testing effects in the section on regression.

### 9.3.2 Using *emmeans* with the Covariate set to Specific Values

In the previous section, we centred a covariate at a specific value in order to test differences between the levels of a within-subject factor at the values of the covariate. We can achieve the same thing by using the *emmeans* subcommand with *compare* and *with(covariate=value)*. For example, the equivalent of the last set of commands in the previous section is:

```
glm test1 test2 with pers1/  
  wsfactor test 2/  
  emmeans=table(test) compare(test) with(pers1=4).
```

If you run these commands, you'll see that the results given in the ANOVA table are not affected: the outcome we're interested in is shown in the *Estimated Marginal Means* section of the output, specifically the pairwise comparisons:

#### Pairwise Comparisons

Measure:MEASURE\_1

(I) test	(J) test	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	-.162	.168	.338	-.496	.172
2	1	.162	.168	.338	-.172	.496

Based on estimated marginal means

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

We can say that, according to the model, when *pers1* is 4, the difference between *test1* and *test2* is .162, which is not significantly different from zero ( $p = .338$ ).

A very useful result also produced by *emmeans* shows the means of *test1* and *test2* at the specified value of the covariate:

#### Estimates

Measure:MEASURE\_1

test	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	4.559 <sup>a</sup>	.124	4.313	4.805
2	4.721 <sup>a</sup>	.124	4.474	4.968

a. Covariates appearing in the model are evaluated at the following values: pers1 Personality score 1 = 4.

These means can be used to produce a plot showing how the differences between the levels of a within-subject effect (*test*, in this case) at different values of a covariate, which allows us to illustrate the nature of an interaction between a within-subject factor and a continuous covariate. The interaction between *test* and *pers1* isn't significant, but let's say we want to illustrate it anyway. We decide to use three values of *pers1*, the mean (4.9) and minus and plus the standard deviation, which is .763, giving 4.137 and 5.663. We carry out the analysis three times (or run one analysis with three different versions of the

*emmeans* subcommand), specifying a different value in *with(pers1=value)* each time. We obtain the following tables:

#### Estimates

Measure:MEASURE\_1

test	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	4.615 <sup>a</sup>	.080	4.456	4.774
2	4.869 <sup>a</sup>	.080	4.710	5.028

a. Covariates appearing in the model are evaluated at the following values: pers1 Personality score 1 = 4.900.

#### Estimates

Measure:MEASURE\_1

test	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	4.568 <sup>a</sup>	.113	4.342	4.793
2	4.744 <sup>a</sup>	.114	4.518	4.969

a. Covariates appearing in the model are evaluated at the following values: pers1 Personality score 1 = 4.137.

#### Estimates

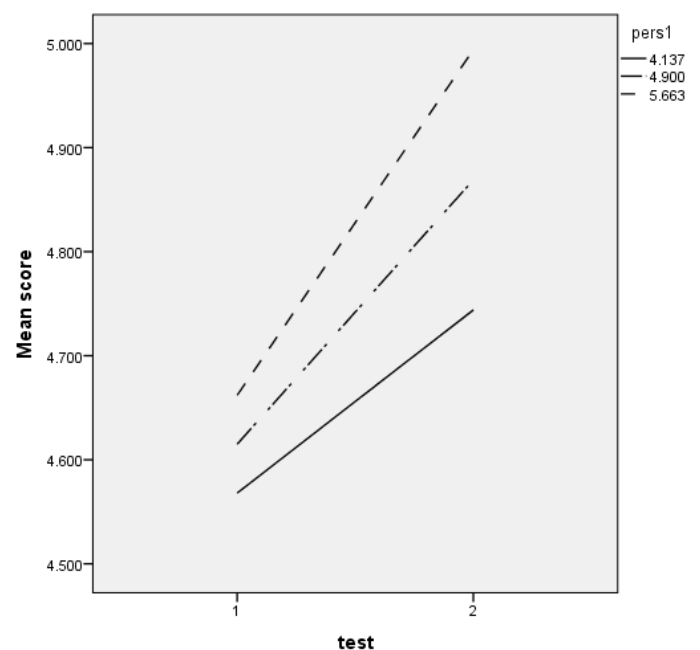
Measure:MEASURE\_1

test	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	4.662 <sup>a</sup>	.114	4.437	4.888
2	4.994 <sup>a</sup>	.114	4.768	5.220

a. Covariates appearing in the model are evaluated at the following values: pers1 Personality score 1 = 5.663.

The means can be entered into an SPSS dataset, and then plotted:

	test	score	pers1
1	1	4.615	4.900
2	2	4.869	4.900
3	1	4.568	4.137
4	2	4.744	4.137
5	1	4.662	5.663
6	2	4.994	5.663
7			



### 9.3.3 A Within-Subject Factor and a Varying Covariate

In Section 9.3, the possibility of an analysis of covariance in which there was a different value of the covariate for each level of the within-subject factor was discussed. When the within-subjects factor is time, this is called a time-varying covariate, but there can be varying covariates for any type of within-subjects factor, as in the illumination example given in Section 9.3.

The question in this of analysis is 'Can the difference(s) between the means for the different levels of the within-subject factor be accounted for by differences in the means of the covariates?'

For this example we'll return to the two variables *test1* and *test3*. The covariate will be represented by two variables, *pers1* and *pers3*. We'll imagine the following scenario: The subjects were measured on two occasions, some time apart, on *test1*, then on *test3*. The test results are known to be affected by the subject's mood at the time of the test, so a 'state' measurement was made at the same time as each of the tests, giving rise to *pers1* and *pers3*. The first question is whether the results for *test1* and *test3* are different. The second question is whether any difference could be accounted for by differences in mood, or less likely, whether mood differences are obscuring differences in test results.

The first question can be answered by carrying out a paired *t*-test or a one-way repeated measures ANOVA. We'll take the easy way out and do two paired *t*-tests, one for *test* and one for *pers*. Click on *Analyze* → *Compare Means* → *Paired-Samples T Test* and select *test1 – test3* and *pers1 – pers3* as *Paired Variables*. (Annoyingly, in point-and-click, it isn't possible to control the order of the variables in each pair, so *test3* is always subtracted from *test1*, rather than *test1* from *test3*, which would make more sense, since a negative difference would mean that the score has gone down from *test1* to *test3*. Fortunately, we can use syntax to get the results in the form we want them.)

#### Syntax

*t-test pairs=test3 pers3 with test1 pers1 (paired).*

Note: Using *with* and *paired* means that only the appropriate pairs of variables will be compared. If the variables were simply listed, every one would be compared with every other one.

The *t*-test results are as follows:

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	TEST3 Test score 3	4.31	99	.855	.086
	TEST1 Test score 1	4.61	99	.794	.080
Pair 2	PERS3 Personality score 3	4.71	99	.794	.080
	PERS1 Personality score 1	4.90	99	.763	.077

**Paired Samples Test**

		Paired Differences		t	df	Sig. (2-tailed)
		Mean	Std. Deviation			
Pair 1	TEST3 Test score 3 - TEST1 Test score 1	-.31	1.091	-2.812	98	.006
Pair 2	PERS3 Personality score 3 - PERS1 Personality score 1	-.19	1.060	-1.762	98	.081

It can be seen that the means for both *test3* and *pers3* are lower than those for *test1* and *pers1* so, although the difference is not significant for *pers*, it may account for some of the difference in *test*. However, there is one more piece of information which is relevant – the correlation between the two difference scores. This can be obtained as follows:

*compute t31diff = test3 - test1.*

*compute p31diff = pers3 - pers1.*

*correlations t31diff p31diff.*

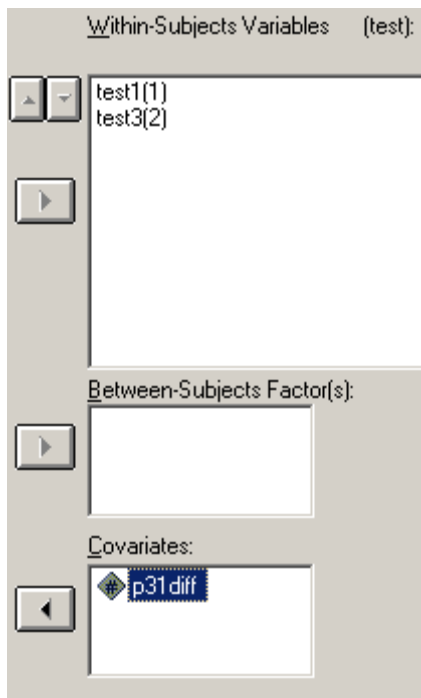
**Correlations**

		T31DIFF	P31DIFF
T31DIFF	Pearson Correlation	1	-.320
	Sig. (2-tailed)	.	.001
	N	99	99
P31DIFF	Pearson Correlation	-.320	1
	Sig. (2-tailed)	.001	.
	N	99	99

This is an interesting result, because the negative correlation means that as the difference between *pers3* and *pers1* becomes more negative (i.e., *pers3* decreases relative to *pers1*), the difference between *test3* and *test1* becomes more positive (i.e.,

*test3* increases relative to *test1*). This means that, far from accounting for the decrease in the score from *test1* to *test3*, the lower score for *pers3* relative to *pers1* is working in the opposite direction to the decrease in the *test* score. We therefore wouldn't expect the analysis of covariance to reduce the effect of *test* – if anything, it may increase it.

The analysis of covariance is carried out by using the difference between *pers3* and *pers1* as the covariate. The syntax above (*compute p31diff = pers3 - pers1.*) created the relevant difference score. Specify the analysis with *Analyze* → *General Linear Model* → *Repeated Measures*, specify the within-subject factor as *test* and enter *test1* and *test3* as the within-subject variables. Enter *p31diff* as a covariate. The display should look like this:



The relevant results are:

#### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
TEST	6.573	1	6.573	12.188	.001
TEST * P31DIFF	5.984	1	5.984	11.097	.001
Error(TEST)	52.309	97	.539		

As predicted, the *test* effect is greater with the covariate than without it: the *p*-value is .001 versus .006 and, more convincingly, the  $\eta^2$  for the analysis with the covariate (.112, not shown in the output) is greater than that obtained without the covariate (.075, also not shown in the output).

A point to note is that, as with the analysis in which a single covariate is used, the analysis is testing the effect of the within-subject factor at the point where the covariate is equal to zero. A value of zero for the difference between the covariates is the point at which the two covariates have equal values, and this is exactly the point at which we want to compare the two *test* variables.

A final point is that the strategy given above works well when the within-subject factor has two levels and there are two covariate measurements, but to cope properly with factors with three or more levels (and a corresponding number of covariate measurements) we have to use the *manova* procedure rather than *GLM*.





## 10. Analysing a Repeated Measures Design with Stacked Data

As mentioned way back in Section 1.4, and also in Section 5, which introduced repeated measures analyses, *GLM* is set up to carry out analyses with data in the multivariate layout, but can be used with 'stacked' data, in which each observation, rather than each subject, is a case in the SPSS data file. The main reason for knowing how to carry out analyses with stacked data is that subjects are not automatically lost when there are missing observations. For example, if subjects are measured on *test1*, *test2* and *test3*, each subject will be represented by three observations in the stacked dataset. If one of the observations is missing, the subject is still represented by the other two observations.

For the purposes of illustration, a stacked dataset will be created from the original *glmdemo.sav* and a mixed analysis will be carried out with *site* as the within-subject factor and *grp* as the between-subject factor. The syntax used to produce the stacked dataset is shown in the box below. Explanatory comments are preceded by asterisks.

### Box 5: The Syntax Used to Produce the Stacked Dataset

\* The 99 cases are sorted by *grp*, then a variable called *swgrp* (short for subject-within-group) is created. At the beginning of the subjects in *grp* 2, the variable is reset so that the value of *swgrp* begins at one for subjects in each group (this is required by *GLM*). *Lag* refers to the subject preceding the current one. *Leave* tells SPSS not to initialise *swgrp* to missing for each new case (which is the default), but to use the number from the preceding case, so that *swgrp* is incremented by one for each new case.

```
sort cases by grp.
compute swgrp=swgrp+1.
if (lag(grp) eq 1 and grp eq 2)swgrp=1.
leave swgrp.
execute.
```

\* Analyses of stacked datasets with large numbers of subjects can be very slow. In order to have a dataset which can be processed reasonably quickly, and which is also balanced, 20 subjects are selected from each *grp*.

```
select if (swgrp le 20).
```

\* The *varstocases* procedure is used to create a dataset containing a variable called *resp* which is made up of the values for *site1*, *site2* and *site3* 'stacked' on top of each other. A variable called *site* with the codes 1, 2 and 3 is also created to show which site each value of *resp* came from. Only the new variables and *id*, *swgrp* and *grp* are kept in the new dataset.

```
varstocases / make resp from site1 site2 site3
/index = site(3)/
/keep = id swgrp grp.
```

\* Note that the *varstocases* command, and also its opposite number, *casestovars*, can be accessed by clicking on *Data/Restructure*.

The data for the first three subjects in the new dataset look like this:

	id	swgrp	grp	site	resp
1	68	1.00	1	1	4.71
2	68	1.00	1	2	4.73
3	68	1.00	1	3	3.54
4	69	2.00	1	1	5.46
5	69	2.00	1	2	4.34
6	69	2.00	1	3	1.56
7	70	3.00	1	1	4.71
8	70	3.00	1	2	4.08
9	70	3.00	1	3	2.02

Unfortunately the analysis of these data cannot be done using point-and-click (if there were only within-subject factors, it could be). The syntax for the mixed model analysis is:

```
glm resp by swgrp site grp/
  random=swgrp/
  design=site grp site*grp swgrp(grp) site*swgrp(grp).
```

Other subcommands could be added, but only the minimum are given for the sake of simplicity. The first point to note is that subject (*swgrp*) is treated as a factor. This is a strategy for dealing with the fact that there are multiple observations for each subject. If we didn't declare subject as a factor, the observations from a given subject would be treated as if they came from different subjects and were therefore independent of each other. Since the responses of a given subject are likely to be correlated with each other, treating them as if they were independent could produce misleading results. Notice also that the *swgrp* factor is declared to be *random* (rather than *fixed*). This means that the levels of the factor (which are subjects) were not decided ahead of time to be at certain fixed values (like the three values of *site*, or the two values of *grp*) but are a random selection from a population of possible values (i.e., subjects). The specification of a random factor in ANOVA has consequences in terms of the error terms which are used to test various effects. For example, a within-subject factor such as *site* is not tested against an overall error term, but against the interaction of that factor and the random factor, in this case *site\*swgrp*.

The second point to note is that the *design* subcommand uses the notation *swgrp(grp)* to tell *GLM* that the *swgrp* factor is *nested* under *grp*. When factor *A* is nested under factor *B*, it means that some levels of *A* occur only under one level of *B* and other levels of *A* occur only under another level of *B*. For example, a piece of educational research may involve giving two groups of schools different levels of funding (*high* and *low*, say) and measuring the result for each school. If Kent Road Public School, which is one level of a *school* factor, is in the *high* funding group, it cannot also be in the *low* funding group, and the same is true for every other school. In this way, the *school* factor is *nested* under the *funding* factor. Back to our example, if a subject is in *grp* = 1, they cannot also be in *grp* = 2: the *grp* and *swgrp* factors are not *crossed*.

The relevant output for the above commands is as follows:

### Tests of Between-Subjects Effects

Dependent Variable: RESP

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	1606.266	1	1606.266	2072.696	.000
	Error	29.449	38	.775 <sup>a</sup>		
SITE	Hypothesis	60.874	2	30.437	49.622	.000
	Error	46.616	76	.613 <sup>b</sup>		
GRP	Hypothesis	5.783E-02	1	5.783E-02	.075	.786
	Error	29.449	38	.775 <sup>a</sup>		
SITE * GRP	Hypothesis	.673	2	.336	.548	.580
	Error	46.616	76	.613 <sup>b</sup>		
SWGRP(GRP)	Hypothesis	29.449	38	.775	1.263	.192
	Error	46.616	76	.613 <sup>b</sup>		
SWGRP * SITE(GRP)	Hypothesis	46.616	76	.613	.	.
	Error	.000	0	. <sup>c</sup>		

a. MS(SWGRP(GRP))

b. MS(SWGRP \* SITE(GRP))

c. MS(Error)

From this table we can see that there is a significant difference among the three sites, but no significant *grp* or *grp\*site* effects. Note that, as mentioned above, the *site* effect is tested against the *site\*swgrp* interaction (called 'SWGRP\*SITE(GRP)' in the table).

Just to reassure ourselves that the two methods give the same results, we can carry out an analysis on the same 40 cases (20 in each *grp*) but which are in the same multivariate layout we have used up to now. The syntax is

```
glm site1 site2 site3 by grp/
wsfactor site 3.
```

and the (sphericity assumed) output is:

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
SITE	60.874	2	30.437	49.622	.000
SITE * GRP	.673	2	.336	.548	.580
Error(SITE)	46.616	76	.613		

**Tests of Between-Subjects Effects**

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1606.266	1	1606.266	2072.696	.000
GRP	5.783E-02	1	5.783E-02	.075	.786
Error	29.449	38	.775		

As can be seen, the sums of squares, mean squares, and the  $F$ - and  $p$ -values are the same as those for the analysis based on the stacked data.

One thing to bear in mind is that with stacked data, it isn't possible to get *GLM* to check for the circularity – in fact, with stacked data, *GLM* is blithely 'unaware' that it is carrying out a repeated measures analysis, and applies the univariate approach without a second thought. If you plan to carry out an analysis with stacked data so as not to lose too many subjects, it might pay to run a preliminary analysis with the data in multivariate layout (even with some subjects missing) to see if there may be a problem with the assumptions.

**10.1 Mixed Linear Models**

SPSS has another procedure for analysing data in stacked format, one which is more sophisticated than *GLM* at dealing with non-independent data (e.g., repeated observations from one person, or observations from numbers of employees from each of a number of organisations). The sophistication of the *mixed* procedure lies in its ability to model different forms of dependency among the observations. We spent a great deal of time worrying about whether repeated measures data met the criteria for *circularity*, but other forms of dependency are possible, and *mixed* allows us to consider them. In addition, *mixed* supplies more accurate estimates of effects than *GLM* when datasets are unbalanced. In the present case, the data are balanced and, with the assumption of circularity, the *mixed* procedure gives identical results to those obtained with *GLM*. Even if the assumptions are relaxed by *mixed* to the point where the variance-covariance matrix is 'unstructured' – anything goes in terms of variances and covariances – the conclusions do not change (this is partly due to the fact they are so clearly significant or non-significant), although the model does fit the data better than one assuming circularity. The *mixed* syntax for the unstructured analysis is as follows:

```
mixed resp by site grp /
  fixed=intercept site grp site*grp /
  repeated=site | subject(id) covtype(un) /
  print=solution testcov.
```

However, a fuller description of the *mixed* procedure is outside the scope of this handbook.

## 11. Carrying Out Regression Analyses with GLM

The *GLM* procedure enables the user to carry out univariate analyses with both categorical and numeric independent variables, so what makes regression analyses different from any other *GLM* analysis with a single dependent variable? I'd argue that there's nothing whatsoever that's unique to regression analyses, but that perhaps it's a useful way to identify analyses with a set of related concerns which differentiate them from the standard ANOVA or ANCOVA. Firstly, the analysis may be concerned with non-experimental data, so that the independent variables are more likely to be correlated, rather than being independent by design, as they are in experiments. Secondly, there may be a concern with the prediction of the outcome variable, the adequacy of the prediction, and the contribution of each independent variable to the prediction, especially in the light of the intercorrelations of the independent variables. Thirdly, following on from the first two concerns, there may be an interest in dropping some 'redundant' independent variables from the analysis and arriving at a 'reduced' model, which contains only independent variables which make a significant unique contribution to predicting the outcome, or which have a unique association with the outcome variable in the sense that they share variance which isn't shared by other variables in the model. Fourthly, there may be an interest on the part of the user in creating his or her own coding for categorical variables, rather than letting the procedure do it behind the scenes. This coding may be done in the interests of interpretation via the signs and values of the regression coefficients, which are a part of the output the ANOVA or ANCOVA user may not find much use for. Fifthly, the thinking involved in regression analysis leads naturally to the use of path analysis and structural equation modelling in a way that the use of ANOVA does not. Finally, there may be an interest in interactions between categorical and numeric variables, and between numeric variables, and it is these interactions which are dealt with in detail here. No attempt is made to cover other topics associated with regression analysis, such as model reduction.

### 11.1 Coding Categorical Variables

This section assumes you've read Sections 1.1 *Factors and Covariates*, and 2.4 *Regression Coefficients*.

To summarise the earlier sections: If a variable is specified as a *Fixed Factor* in point-and-click, or after the *by* in syntax, *GLM* assumes it is a categorical variable and presents results based on its own codings of the variable. The results given in the ANOVA table are based on what is called *effect* or *deviation* coding. This kind of coding is based on codes of (-1, 0 and 1) and the regression coefficients produced by such coding show how the mean of the dependent variable for each category differs from the overall mean of the dependent variable. The results given in the table of parameter estimates, on the other hand, are for *dummy* coding, which is based on (0,1) codes. The coefficients for a given category show how the mean of the dependent variable for that category differs from the mean for the reference category, which is always the highest-numbered category.

In this section on regression, we're generally going to avoid the issues arising from the codings used by *GLM* by not specifying *Fixed Factor* and by not using *by*. Most of our independent variables are going to be specified as *Covariates*, or after *with*. That way, we'll control the coding and make sure that the regression coefficients are easy to interpret. Mostly, we'll use dummy coding, rather than the various alternatives (which are

sometimes very useful). We will, however, sometimes ask *GLM* to treat categorical variables as fixed factors in order to get access to pairwise tests and graphs of adjusted means. That way, we can have the best of both worlds.

## 11.2 The Equivalence of Fixed Factor (*by*) and Covariate (*with*) for a Single Dummy Variable

If an independent variable has only two categories, as do *gender* or *grp* in the present dataset, a perfectly valid result will be obtained regardless of whether it is entered as a *Fixed Factor* (after *by*) or as a *Covariate* (after *with*). This will be demonstrated with a dependent variable called *tc*, which is the average of the variables *typ1con1* to *typ2con2*, and the dichotomous grouping variable, *grp*, as the independent variable. For the purposes of this and subsequent analyses, a dummy-coded (0,1) version of *grp* will be used, although the points made in Section 11.2 apply even if the original (1,2) coding had been used. The *tc* variable and the recoded *grp* (*grpd*) are created, and the mean of *tc* by *grpd* is obtained, as follows:

```
compute tc=mean(typ1con1 to typ2con2).
recode grp (1=0)(2=1) into grpd.
print format grpd (f1).
value labels grpd 0 'Control' 1 'Experimental'.
means tc by grpd.
```

### Report

TC			
GRPD	Mean	N	Std. Deviation
0 Control	4.4957	51	.78780
1 Experimental	4.7932	48	.85355
Total	4.6399	99	.82969

The tables of means show that the control and experimental means differ by  $4.7932 - 4.4957 = .2975$ , or  $4.4957 - 4.7932 = -.2975$ , depending which is subtracted from the other.

First, we'll specify *grpd* as a *Fixed Factor* (after *by*).

Point-and-click: *Analyze* → *General Linear Model* → *Univariate*. Specify *tc* as the *Dependent Variable*, *grpd* as a *Fixed Factor*, and check *Parameter estimates* as an *Option*.

Syntax: *glm tc by grpd/print=parameters*.

The output is as follows:

### Between-Subjects Factors

	Value Label	N
GRPD 0	Control	51
1	Experimental	48

**Tests of Between-Subjects Effects**

Dependent Variable: TC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.188 <sup>a</sup>	1	2.188	3.252	.074
Intercept	2133.520	1	2133.520	3170.537	.000
GRPD	2.188	1	2.188	3.252	.074
Error	65.273	97	.673		
Total	2198.800	99			
Corrected Total	67.462	98			

a. R Squared = .032 (Adjusted R Squared = .022)

**Parameter Estimates**

Dependent Variable: TC

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.793	.118	40.482	.000	4.558	5.028
[GRPD=0]	-.2975	.165	-1.803	.074	-.625	2.992E-02
[GRPD=1]	0 <sup>a</sup>	.	.	.	.	.

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

The main result, which is reported in both the ANOVA table and the Parameter Estimates table, is that the *p*-value for the significance test is .074. Note that, as discussed previously, *GLM* bases the parameter estimates on a dummy-coded model with the highest-numbered category (1 in this case) as the reference category. Thus it makes perfect sense that the parameter estimate for *grpd* is -.2975, the result of subtracting the mean for the reference category (the experimental group) from the mean for the lowest-numbered category (the control group).

Before we try out the *Covariate/with* option, let's check out what's happening with the intercept. On this, the ANOVA table and the Parameter Estimates table disagree: the latter gives the coefficient as 4.793 and the *t*-value as 40.482. The square of 40.482 is 1638.8, not 3170.5, as shown in the ANOVA table, so evidently two different things are being tested. At least we can be clear on the meaning of the result given in the Parameter Estimates table: the intercept shows the value of *tc* when *grpd* is zero, and we know that the experimental group was coded zero by *GLM*, and that the value of 4.793 is the mean for that group, as the table of means shows. So what's being tested in the ANOVA table? The answer can be summarised as follows:

- For the results used in the ANOVA table, *GLM* uses effect or deviation coding for *grpd*, in which 1 is coded -1 and 0 is coded 1.
- This form of coding results in a regression coefficient for *grpd* (not given in our results) which shows the difference between the mean for the group coded 1 and the unweighted overall mean (this is a whole topic in itself, but it won't be pursued here).
- Given that the experimental group is coded as -1 and the control group as 1, what does the intercept, which shows the value of *tc* when *grpd* is zero, mean?

As you might expect, since zero is midway between -1 and 1, the intercept shows the overall average of *tc*: more exactly, the unweighted mean of the means of *tc* for the two groups,  $(4.4957 + 4.7932)/2 = 4.6444$ . This value is not shown in the output, but the *F* of 3170.5 is the result of testing whether this value is different from zero.

We'll now return to the main issue and specify *grp* as a *Covariate* (after *with*).

**Point-and-click:** *Analyze* → *General Linear Model* → *Univariate*. Specify *tc* as the *Dependent Variable*, *grp* as a *Covariate*, and check *Parameter estimates* as an *Option*.

**Syntax:** *glm tc with grp/print=parameters*.

The results are shown below. The obvious point is that once again the *p*-value for *grp* is .074 in both the ANOVA table and the Parameter Estimates table, which confirms the assertion made at the beginning of this section about the equivalence of the results obtained for single dummy variables entered as fixed factors or covariates. However, it's also worth noting that, because *GLM* has used the codes we specified, rather than creating its own for the purposes of the results in the ANOVA table, all details of the two tables are identical. We know the intercept is identical because the *F*-ratio in the ANOVA table, 1531.77 is the square of the *t*-value shown in Parameter Estimates table. Finally, note that the coefficient for *grp* shown in the Parameter Estimates table is now positive rather than negative, because our coding made the control group the reference category.

#### Tests of Between-Subjects Effects

Dependent Variable: TC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2.188 <sup>a</sup>	1	2.188	3.252	.074
Intercept	1030.759	1	1030.759	1531.770	.000
GRPD	2.188	1	2.188	3.252	.074
Error	65.273	97	.673		
Total	2198.800	99			
Corrected Total	67.462	98			

a. R Squared = .032 (Adjusted R Squared = .022)

#### Parameter Estimates

Dependent Variable: TC

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.496	.115	39.138	.000	4.268	4.724
GRPD	.297	.165	1.803	.074	-2.992E-02	.625

## 12. A Multiple Regression Analysis with GLM

In this example, *tc* is again the dependent variable, and we are interested in examining the effects of *gender*, *age*, a personality measure, *pers3*, and a grouping variable with four categories (*group*), on *tc*. A researcher would have a rationale for an analysis involving



these particular variables, and also one or more preliminary analyses would have been performed. Because the aim of the example is to demonstrate the use of *GLM* and the interpretation of results, we'll take these steps for granted. Some descriptive statistics for the variables are produced by the following commands, but you're encouraged to carry out your own bivariate analyses so that you can compare the (unadjusted) results with those obtained in the regression analysis.

*descriptives tc age pers3.*  
*frequencies gender group.*

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
TC	99	2.50	6.85	4.6399	.82969
AGE	99	15.00	50.00	24.7576	7.59040
PERS3 Personality score 3	99	3	7	4.71	.794
Valid N (listwise)	99				

GENDER

	Frequency	Percent
Valid 1 male	57	57.6
2 female	42	42.4
Total	99	100.0

GROUP

	Frequency	Percent
Valid 1 Control	29	29.3
2 25 mg/kg	24	24.2
3 50 mg/kg	31	31.3
4 100 mg/kg	15	15.2
Total	99	100.0

Dummy-coded versions of *gender* and *group* are produced as follows:

*recode gender (1=1)(2=0) into gend.*  
*print format gend (f1).*  
*value labels gend 0 'Female' 1 'Male'.*

*recode group (2=1)(missing=sysmis)(else=0) into grp2.*  
*recode group (3=1)(missing=sysmis)(else=0) into grp3.*  
*recode group (4=1)(missing=sysmis)(else=0) into grp4.*

We're now ready to perform the analysis.

Click on *Analyze* → *General Linear Model* → *Univariate*. Specify *tc* as the *Dependent Variable*, and *gend*, *age*, *pers3* and *grp2* to *grp4* as *Covariates*, and check *Parameter estimates* as an *Option*. (Be sure not to enter the independent variables as *Fixed Factors* or *Random Factors*.)

Syntax

```
glm tc with gend age pers3 grp2 grp3 grp4/
print=parameters.
```

The results are shown in the ANOVA and Parameter Estimates tables below. The following points can be made.

- Because the default Type 3 sums of squares were used, the significance tests for each variable in the ANOVA and Parameter Estimates tables are identical. Each variable is tested as if it were fitted last in a sequential analysis. Thus, apart from showing the variance accounted for by the full set of predictors, the ANOVA table doesn't tell us anything that the Parameter Estimates table doesn't, and we'll concentrate on the latter.
- The intercept estimate tells us the estimated value of *tc* when all the independent variables have a value of zero. This would be quite meaningful for the dummy variables, since values of zero stand for groups of subjects but, as can be seen from the descriptive statistics, zero is outside the range of values observed for the numeric variables *age* and *pers3*.

#### Tests of Between-Subjects Effects

Dependent Variable: TC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	25.662 <sup>a</sup>	6	4.277	9.414	.000
Intercept	17.102	1	17.102	37.641	.000
GEND	.196	1	.196	.432	.513
AGE	2.485	1	2.485	5.469	.022
PERS3	1.475	1	1.475	3.246	.075
GRP2	.626	1	.626	1.377	.244
GRP3	2.988	1	2.988	6.576	.012
GRP4	19.418	1	19.418	42.739	.000
Error	41.800	92	.454		
Total	2198.800	99			
Corrected Total	67.462	98			

a. R Squared = .380 (Adjusted R Squared = .340)

#### Parameter Estimates

Dependent Variable: TC

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	2.991	.488	6.135	.000	2.023	3.959
GEND	-.092	.140	-.657	.513	-.369	.185
AGE	.021	.009	2.338	.022	.003	.039
PERS3	.162	.090	1.802	.075	-.017	.340
GRP2	.220	.187	1.173	.244	-.152	.592
GRP3	.471	.184	2.564	.012	.106	.835
GRP4	1.422	.218	6.538	.000	.990	1.854

- The remaining regression coefficients tell us the effect on the estimated value of *tc* of a one-unit increase in each of the independent variables. For example, since *gend* is coded zero for females and one for males, the coefficient for *gend* of *-.092* tells us that, on average, males scored *.092* lower than females (adjusted for the other variables), although this difference is not significant. The value of *.021* for *age* tells us that as *age*

increases by one year, the estimate of *tc* increases by .021 (adjusted for the other variables). This effect is significant.

- It's clear that the *tc* scores differ significantly over group. Since we chose to represent *group* by our own dummy variables *grp2*, *grp3* and *grp4*, we can't tell *GLM* that together they represent *group*, so *GLM* hasn't made an overall test of that factor. However, we can do our own test by saying that if one or more of the three dummy variables is significant at  $.05/3 = .017$ , then the overall effect is significant. Two of the dummy variables for *group* are significant at this level. The coefficient for *grp3* tells us that the estimated value of *tc* for *grp3* is .471 higher than that for the reference group (the group numbered 1 on the original *group* variable, from which the dummy variables were derived), while the estimated mean for *grp4* is 1.422 higher than the mean for the reference group.

**Box 6: Specifying *group* as a *Fixed Factor* to get pairwise tests and graphs of adjusted means**

As mentioned in Section 11.1, even in a regression analysis, it is sometimes very useful to ask *GLM* to treat a categorical variable as a *Fixed Factor* in order to get access to *GLM* facilities which aren't available when variables are specified as covariates. This is such an occasion: we'd like to make all pairwise comparisons of the means of the four groups, and have a graph of means, in both cases adjusted for the other variables in the model.

With point-and-click, remove *grp2*, *grp3* and *grp4* from the *Covariates* slot, and add *group* to the *Fixed Factors* slot. Click on *Plots* and select *group* for the *Horizontal Axis*. Click on *Options* and select *group* in the *Display Means for* slot. Check the *Compare main effects* box.

Syntax:

```
glm tc by group with gend age pers3/  
  print=parameters/  
  emmeans=table(group) compare(group)/  
  plot=profile(group).
```

The results are shown in the tables below:

**Tests of Between-Subjects Effects**

Dependent Variable: TC

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	25.662 <sup>a</sup>	6	4.277	9.414	.000
Intercept	23.713	1	23.713	52.192	.000
GEND	.196	1	.196	.432	.513
AGE	2.485	1	2.485	5.469	.022
PERS3	1.475	1	1.475	3.246	.075
GROUP	20.608	3	6.869	15.119	.000
Error	41.800	92	.454		
Total	2198.800	99			
Corrected Total	67.462	98			

a. R Squared = .380 (Adjusted R Squared = .340)

The ANOVA table gives the results of an overall test of *group*, which is useful for quoting in a write-up. As would be expected, the results for the other variables are identical to those in the original ANOVA table. (The result for the intercept is different, though. If you're interested in finding out why, check out Section 11.2.)

**Parameter Estimates**

Dependent Variable: TC

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.413	.512	8.617	.000	3.396	5.431
GEND	-.092	.140	-.657	.513	-.369	.185
AGE	.021	.009	2.338	.022	.003	.039
PERS3	.162	.090	1.802	.075	-.017	.340
[GROUP=1]	-1.422	.218	-6.538	.000	-1.854	-.990
[GROUP=2]	-1.202	.223	-5.399	.000	-1.645	-.760
[GROUP=3]	-.952	.215	-4.435	.000	-1.378	-.525
[GROUP=4]	0 <sup>a</sup>	.	.	.	.	.

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

Because *GLM* did its own coding for *group*, the highest-numbered group was used as the reference category for *group*, and the coefficients in the Parameter Estimates table reflect this. The results for the other variables are again identical to those in the original analysis.

The remaining tables show the estimated marginal means (which are adjusted for the other variables in the model), and pairwise comparisons and a plot of the means.

Dependent Variable: TC

	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1 Control	4.224 <sup>a</sup>	.128	3.969	4.478
2 25 mg/kg	4.444 <sup>a</sup>	.139	4.168	4.719
3 50 mg/kg	4.694 <sup>a</sup>	.125	4.445	4.944
4 100 mg/kg	5.646 <sup>a</sup>	.175	5.299	5.993

a. Evaluated at covariates appeared in the model: GEND = .58,  
AGE = 24.7576, PERS3 Personality score 3 = 4.71.

**Pairwise Comparisons**

Dependent Variable: TC

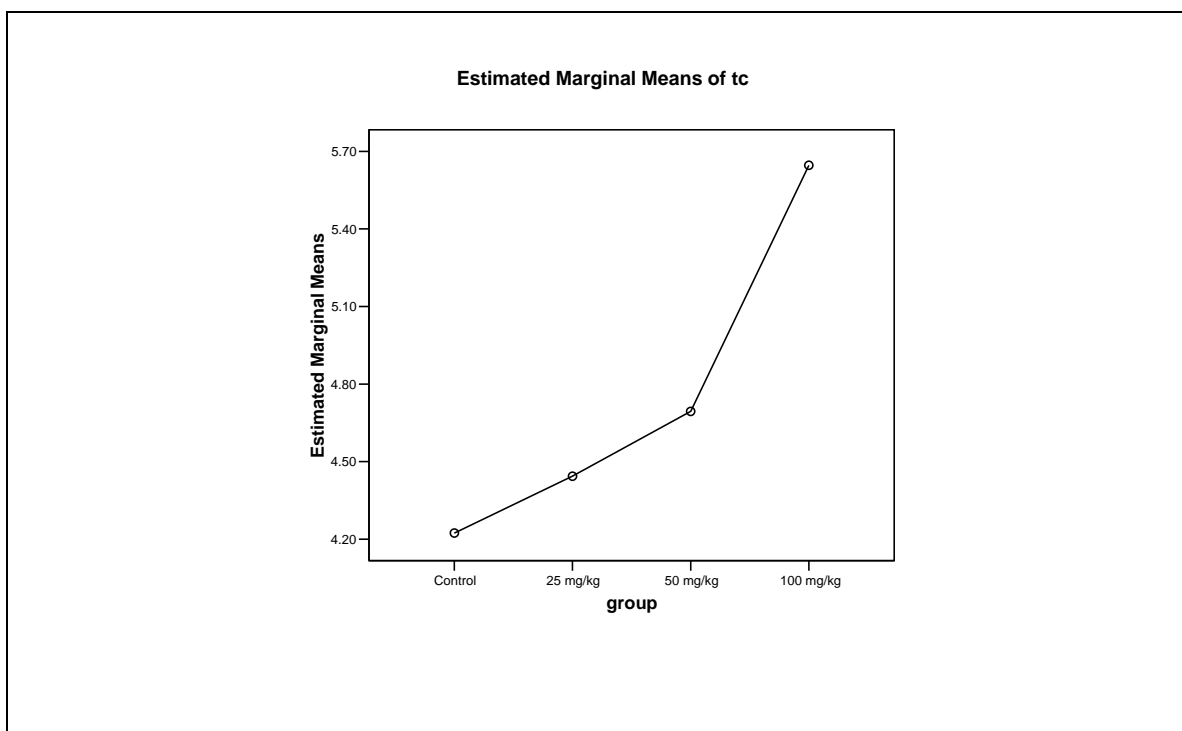
(I)	(J)	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1 Control	2 25 mg/kg	-.220	.187	.244	-.592	.152
	3 50 mg/kg	-.471*	.184	.012	-.835	-.106
	4 100 mg/kg	-1.422*	.218	.000	-1.854	-.990
2 25 mg/kg	1 Control	.220	.187	.244	-.152	.592
	3 50 mg/kg	-.251	.190	.190	-.628	.127
	4 100 mg/kg	-1.202*	.223	.000	-1.645	-.760
3 50 mg/kg	1 Control	.471*	.184	.012	.106	.835
	2 25 mg/kg	.251	.190	.190	-.127	.628
	4 100 mg/kg	-.952*	.215	.000	-1.378	-.525
4 100 mg/kg	1 Control	1.422*	.218	.000	.990	1.854
	2 25 mg/kg	1.202*	.223	.000	.760	1.645
	3 50 mg/kg	.952*	.215	.000	.525	1.378

Based on estimated marginal means

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

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

**Note:** For further information about pairwise comparisons using estimated marginal means, see Section 2.3.3.



### 13. Interactions Among Categorical Variables

In the *glimdemo.sav* dataset, there is a variable *score*, which is an (imaginary) neuropsychological test result. There are also two dummy-coded two-category variables *grpd* (control and experimental) and *brief*. *Brief* shows whether or not subjects received a special briefing ('1' = yes, '0' = no) before they underwent the neuropsychological test under control or experimental conditions. Descriptive statistics are given below.

*means score by grpd by brief.*

#### Report

SCORE				
GRPD	BRIEF	Mean	N	Std. Deviation
0 Control	0 No briefing	12.4968	28	5.73801
	1 Briefing	13.3542	23	5.80822
	Total	12.8835	51	5.72788
1 Experimental	0 No briefing	11.2072	21	4.02769
	1 Briefing	21.5871	27	8.10989
	Total	17.0459	48	8.38843

Most likely, a 2 x 2 independent ANOVA would be used to test the main effects of *grpd*, *brief* and the interaction between the two factors. Section 4 gave details of how to carry out such an analysis by specifying *grpd* and *brief* as fixed factors, and how to carry out various follow-up tests. The required analysis could also be done in the regression context described in Section 11, and this will be outlined below. The main reason for describing this way of doing the analysis is to make some points about the interpretation of interactions. In general, for this analysis, I can't see any advantages in the regression

approach as against the conventional ANOVA approach in *GLM*, which allows the easy plotting of graphs, and the testing of simple and interaction effects.

A main effects model can be tested as follows:

**Point-and-click:** *Analyze* → *General Linear Model* → *Univariate*. Specify *score* as the *Dependent Variable*, *grp* and *brief* as *Covariates*, and check *Parameter estimates* as an *Option*.

**Syntax:** *glm score with grp brief/print=parameters.*

The tables below show that both *grp* and *brief* are significant. The regression coefficient for *grp* suggests that, averaged over the two levels of *brief*, the mean score for subjects in the experimental group was 3.554 higher than that for subjects in the control group. Notice that this difference is not equal to the difference between either the weighted or unweighted means for the two groups (these can be calculated from the table of means above). This is because the main effects model we're testing isn't able to take into account variations in means other than those which are due to the main effects. Because there is no interaction term, variations due to combinations of conditions (e.g., the part of the experimental group which had no briefing) are not taken into account. The extent to which such a reduced model is satisfactory depends on the magnitude of the interaction effect, which we'll test next. Similar comments can be made about the main effect of *brief*.

#### Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1156.977 <sup>a</sup>	2	578.489	13.163	.000
Intercept	3888.379	1	3888.379	88.476	.000
GRPD	308.367	1	308.367	7.017	.009
BRIEF	728.559	1	728.559	16.578	.000
Error	4219.059	96	43.949		
Total	27359.778	99			
Corrected Total	5376.037	98			

a. R Squared = .215 (Adjusted R Squared = .199)

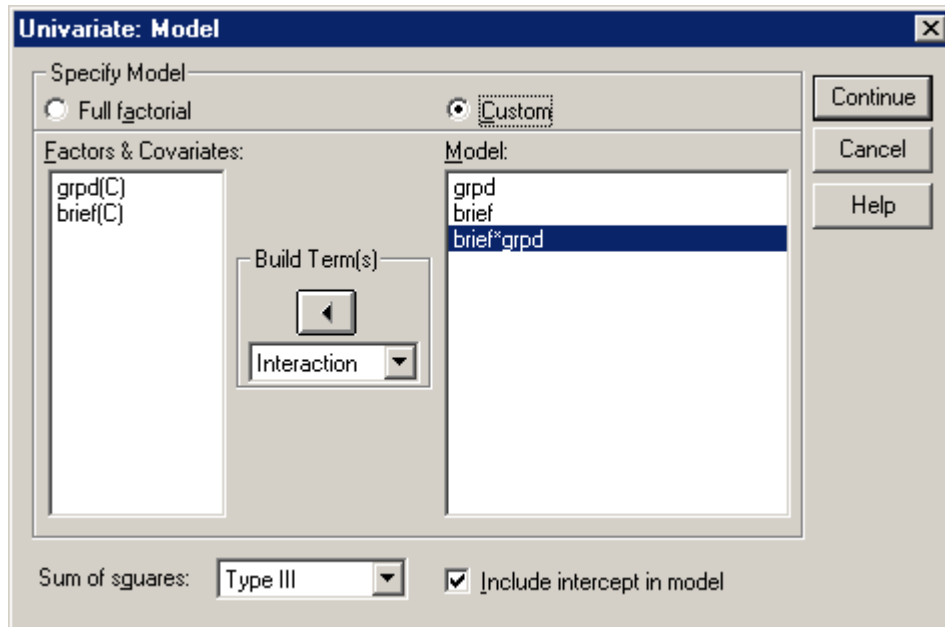
#### Parameter Estimates

Dependent Variable: SCORE

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	10.421	1.108	9.406	.000	8.222	12.620
GRPD	3.554	1.342	2.649	.009	.891	6.216
BRIEF	5.460	1.341	4.072	.000	2.798	8.122

While the default in *GLM* is to test interactions between *Fixed Factors*, the opposite is true for *Covariates*, so we must ask *GLM* to create the *grp* by *brief* interaction term.

Point-and-click: *Analyze* → *General Linear Model* → *Univariate*. Specify *score* as the *Dependent Variable*, *grp**d* and *brief* as *Covariates*, and check *Parameter estimates* as an *Option*. Now the new bit: Click on the *Model* button, then click on the *Custom* radio button. Select both *grp**d* and *brief*, select *Main Effects* in the *Build Terms* slot, and click on the arrow. Repeat, but this time select *Interaction* in the *Build Terms* slot. The display should look like this:



#### Syntax

```
glm score with grp d brief /
  print=parameters /
  design=grp d brief grp d*brief.
```

The *design* subcommand allows the specification of non-default models. Asterisks are used to specify interactions between variables.

#### Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1710.415 <sup>a</sup>	3	570.138	14.776	.000
Intercept	4372.780	1	4372.780	113.327	.000
GRPD	19.959	1	19.959	.517	.474
BRIEF	9.281	1	9.281	.241	.625
GRPD * BRIEF	553.438	1	553.438	14.343	.000
Error	3665.621	95	38.585		
Total	27359.778	99			
Corrected Total	5376.037	98			

a. R Squared = .318 (Adjusted R Squared = .297)



**Parameter Estimates**

Dependent Variable: SCORE

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	12.497	1.174	10.646	.000	10.166	14.827
GRPD	-1.290	1.793	-.719	.474	-4.850	2.270
BRIEF	.857	1.748	.490	.625	-2.613	4.328
GRPD * BRIEF	9.523	2.514	3.787	.000	4.531	14.514

Clearly the interaction is significant! This gives us the excuse to investigate the regression coefficients to see what they can tell us.

The following three principles are crucial to the interpretation of interactions in regression analyses and should come into your head automatically whenever you want to interpret the results of such analyses. We'll state the principles, then use them to interpret the above table. It will also help to look at the interaction graph below, which is annotated with the means.

**Box 7: Three Important Principles**

First a familiar rule:

Rule 1. The coefficient for the intercept shows the estimated value of the dependent variable when all the predictors have values of zero.

Now some new rules:

In a regression equation

$$y = B_0 + B_1 * x + B_2 * z + B_3 * x * z \quad (\text{remember } * \text{ is the multiplication symbol})$$

which includes an interaction term,

Rule 2.

$B_1$  shows the effect of  $x$  on  $y$  when  $z$  is zero, and  
 $B_2$  shows the effect of  $z$  on  $y$  when  $x$  is zero.

Rule 3.

$B_3$  shows how the effect of  $x$  on  $y$  changes with a one-unit increase in  $z$ , OR  
how the effect of  $z$  on  $y$  changes with a one-unit increase in  $x$ .

In the above, 'the effect of  $x$  on  $y$ ' is the amount  $y$  increases with a one-unit increase in  $x$ , which is shown by the regression coefficient; it is the slope of the relevant regression line.

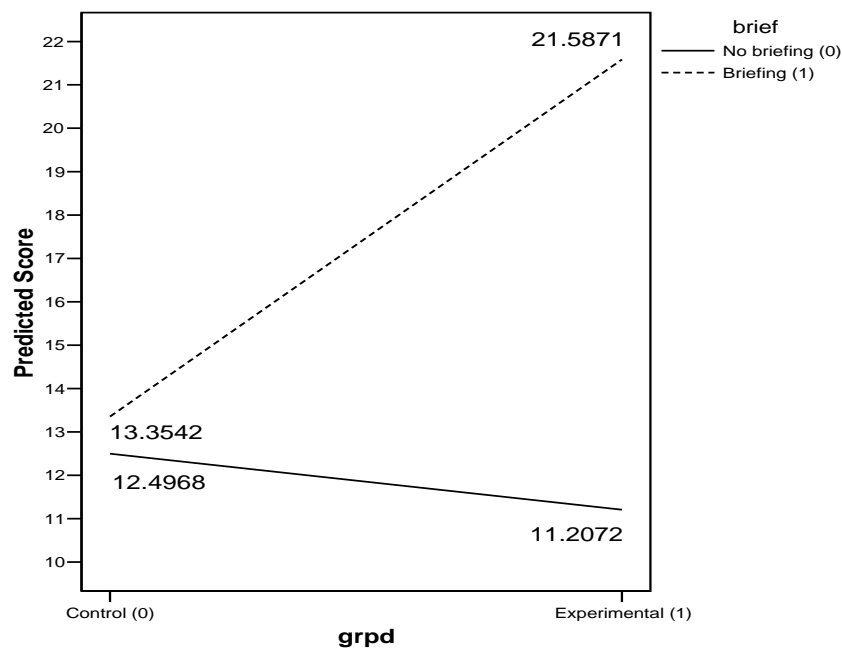
Armed with the above principles, we'll now look at the table of Parameter Estimates.

**Rule 1:** The coefficient of the intercept is **12.497**. According to Rule 1, this coefficient shows the estimated value of *score* when both *grp* and *brief* are zero. As can be seen in the table of means at the beginning of this section, that is the Control/No Briefing group, and the mean of *score* for that group is indeed **12.497**.

**Rule 2:** The coefficient for *grp*, **-1.29**, shows the effect of *grp* when *brief* is zero. This is the effect of a one-unit increase in *grp*, from *grp* equals 0 to *grp* equals 1, for the no briefing groups. The relevant means are 12.4968 (*grp* = 0, *brief* = 0) and 11.2072 (*grp* = 1, *brief* = 0). To see the effect of a one-unit increase in *grp* we need to subtract the value for *grp* = 0 from the value for *grp* = 1, which is  $11.2072 - 12.4968 = -1.2986$ .

A similar line of reasoning occurs for the coefficient for *brief*, **.857**, which shows the effect of *brief* when *grp* is zero. This is the effect of a one-unit increase in *brief*, from *brief* equals 0 to *brief* equals 1, for the control groups. The relevant means are 12.4968 (*brief* = 0, *grp* = 0) and 13.3542 (*brief* = 1, *grp* = 0). To see the effect of a one-unit increase in *brief* we need to subtract the value for *brief* = 0 from the value for *brief* = 1, which is  $13.3542 - 12.4968 = .8574$ .

**Note:** If you have trouble deciding which value to subtract from the other in the above sorts of calculations, think of it this way: you want to know how much a dependent variable changes with a one-unit increase in an independent variable. Here's an analogy. If we know that Penrith is one hour's drive from Sydney, and Katoomba is three hour's drive from Sydney, we would subtract the time taken to get to the closer location (Penrith) from the time taken to the further location (Katoomba) to work out the time it takes to get from Penrith to Katoomba. If we did the subtraction the other way around, we'd find that it took -2 hours to drive from Penrith to Katoomba. (Any relationship to actual driving times is co-incidental.)



Rule 3: The regression coefficient for the *grp**d* \* *brief* is **9.523**.

(1) This is the result of a comparison between the effect of *grp**d* when *brief* = 0 and the effect of *grp**d* when *brief* = 1.

The effect of *grp**d* when *brief* = 0:  $11.2072 - 12.4968 = -1.2896$

The effect of *grp**d* when *brief* = 1:  $21.5871 - 13.3542 = 8.2329$

The difference between the effect of *grp**d* when *brief* = 0 and the effect of *grp**d* when *brief* = 1 (remembering to subtract the result for *brief* = 0 from the effect for *brief* = 1):

$$8.2329 - (-1.2896) = \mathbf{9.5225}$$

In one bite:  $(21.5871 - 13.3542) - (11.2072 - 12.4968) = \mathbf{9.5225}$

(2) This is equivalently the result of a comparison between the effect of *brief* when *grp**d* = 0 and the effect of *brief* when *grp**d* = 1.

The effect of *brief* when *grp**d* = 0:  $13.3542 - 12.4968 = .8574$

The effect of *brief* when *grp**d* = 1:  $21.5871 - 11.2072 = 10.3799$

The difference between the effect of *brief* when *grp**d* = 0 and the effect of *brief* when *grp**d* = 1 (remembering to subtract the result for *grp**d* = 0 from the effect for *grp**d* = 1):

$$10.3799 - .8574 = \mathbf{9.5225}$$

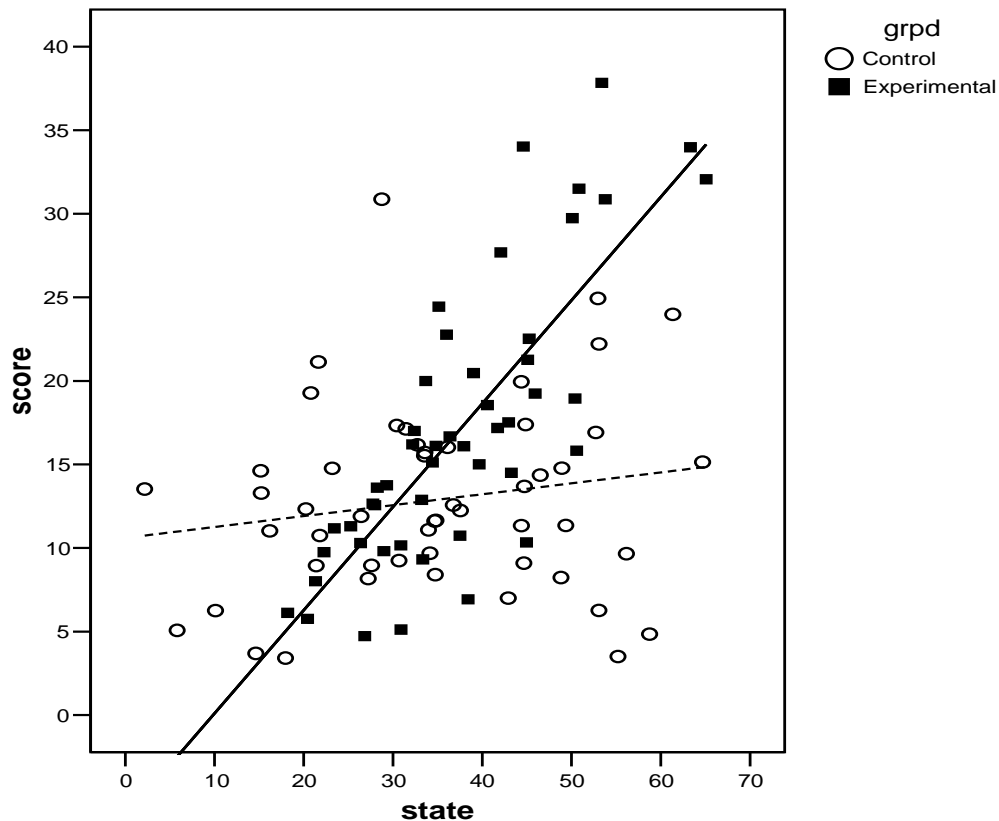
In one bite:  $(21.5871 - 11.2072) - (13.3542 - 12.4968) = \mathbf{9.5225}$

The preceding section is crucial to the understanding of interactions, not just with dummy coding, but also with other coding schemes. If you understand what we've just done, you'll find the remaining sections, which deal with interactions involving numeric variables, make a great deal of sense.

The methods used in the preceding section generalise easily to the case in which more than one dummy variable is needed to represent a factor (such as *group* in *glmdemo.sav*). When interpreting the results in such cases, the main thing is to examine one dummy variable at a time and consider the relationship between the category it represents and the reference category before going on to the next dummy variable.

## 14. The Interaction Between a Categorical Variable and a Numeric Variable

Staying with the neuropsychological measure, *score*, as the dependent variable, and *grp**d* (control versus experimental) as a grouping variable, we now consider an additional independent variable, a numeric variable called *state*. *State* is a measure of a 'state' trait which is hypothesised to affect the performance (as measured by *score*) of subjects in the experimental group, but to have no particular effect on the performance of subjects in the control group. In other words, an interaction between *grp**d* and *state* is hypothesised. A graph with the regression of *score* on *state* fitted separately for each group, shown below, certainly suggests the presence of an interaction, with subjects in the experimental group who have higher values for the *state* variable having higher values of *score*.



The following commands allow the test of a model containing the interaction term.

Point-and-click: *Analyze* → *General Linear Model* → *Univariate*. Specify *score* as the *Dependent Variable*, *grp*d and *state* as *Covariates*, and check *Parameter estimates* as an *Option*. Click on the *Model* button, then click on the *Custom* radio button. Select both *grp*d and *state*, select *Main Effects* in the *Build Terms* slot, and click on the arrow. Repeat, but this time select *Interaction* in the *Build Terms* slot.

Syntax

```
glm score with grp d state/
print=parameters/
design=grp d state grp d*state.
```

This time we'll consider the regression coefficients only.

#### Parameter Estimates

Dependent Variable: SCORE

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	10.605	1.947	5.448	.000	6.741	14.470
GRPD	-16.686	3.484	-4.790	.000	-23.602	-9.771
STATE	.065	.051	1.273	.206	-3.657E-02	.167
GRPD * STATE	.553	.090	6.123	.000	.374	.732

Firstly, the interaction is clearly significant. The coefficient indicates that as the value of

*grp* goes from 0 to 1, the effect of *state* on *score* increases by .553. (A rough check on this figure from inspection of the graph is that  $(37 - 0)/(70 - 10) - (15 - 11)/(70 - 0) = .560$ , which is close to .553.)

Now, let's look at the coefficients for the two variables. First, remembering the second of the crucial principles described above, the value of .065 says that when *grp* is 0 (the control group), the slope of the regression line linking *score* with *state* is only .065, and is not significantly different from zero ( $t(95) = 1.273, p = .206$ ). This is certainly consistent with what appears in the graph  $[(15 - 11)/(70 - 0) = .057]$ . What about the coefficient of -16.686 for *grp*? This is the difference between the estimated means for the control and experimental groups when *state* equals zero. As can be seen, zero is 'off the scale' for *state*; so a comparison between groups at zero is not very useful.

We'll now look at how we can take advantage of the second principle above to (i) obtain further information, and (ii) obtain more useful regression coefficients.

(i) It was certainly useful to be able to say that the effect of *state* on *score* is not significantly different from zero for the control group. It would also be useful to know the effect of *state* on *score* and whether it is different from zero for the experimental group. We can easily find this out by temporarily reversing the coding of *grp*, so that now the experimental group is coded zero. The coding means that the regression coefficient for *state* now shows the relationship between *state* and *score* for the experimental group. The following commands will obtain this result (although the reversal of the code for *grp* can be done with point-and-click, it is shown in syntax form for brevity):

```
temporary.
recode grp (0=1)(1=0).
glm score with grp state /
  print=parameters /
  design=grp state grp*state.
```

The resulting regression coefficients are:

**Parameter Estimates**

Dependent Variable: SCORE

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	-6.081	2.889	-2.105	.038	-11.816	-.346
GRPD	16.686	3.484	4.790	.000	9.771	23.602
STATE	.618	.074	8.321	.000	.471	.766
GRPD * STATE	-.553	.090	-6.123	.000	-.732	-.374

The coefficient for *state* shows that the slope of the regression line for the experimental group is .618, and is significantly different from zero. The value of .618 checks out with a rough calculation based on the graph:  $(37 - 0)/(70 - 10) = .617$ .

(ii) The mean of *state* is:

	N	Minimum	Maximum	Mean	Std. Deviation
STATE	99	2.18	65.09	36.1113	13.12891

*descriptives state.*

If 36.1113 is subtracted from *state* for each subject, the mean of *state* will now be zero, so when the *centred* variable is entered in the regression, the coefficient for *grpd* will show its effect at the mean of *state*.

*temporary.*

*compute state=state – 36.1113.*

*glm score with grpd state/*

*print=parameters/*

*design=grpd state grpd\*state.*

The table below gives a value of 3.274 for *grpd*. Referring to the graph and taking a line up from 36 on the *x*-axis, it does appear that the lines are about 3 apart. We can say that, at the mean of *state*, the mean of the experimental group is significantly higher than that for the control group.

### Box 8: Why Stop at the Mean?

The size and significance of the difference between the two groups can be obtained for any value of *state* simply by subtracting the appropriate value from *state* before the analysis is run. Of course, this technique is only sensible if the subtracted value is well within the range of the values of *state* actually observed.

Parameter Estimates

Dependent Variable: SCORE

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	12.964	.767	16.894	.000	11.440	14.487
GRPD	3.274	1.104	2.965	.004	1.082	5.467
STATE	.065	.051	1.273	.206	-3.657E-02	.167
GRPD * STATE	.553	.090	6.123	.000	.374	.732

#### 14.1 Producing a Graph of the Interaction

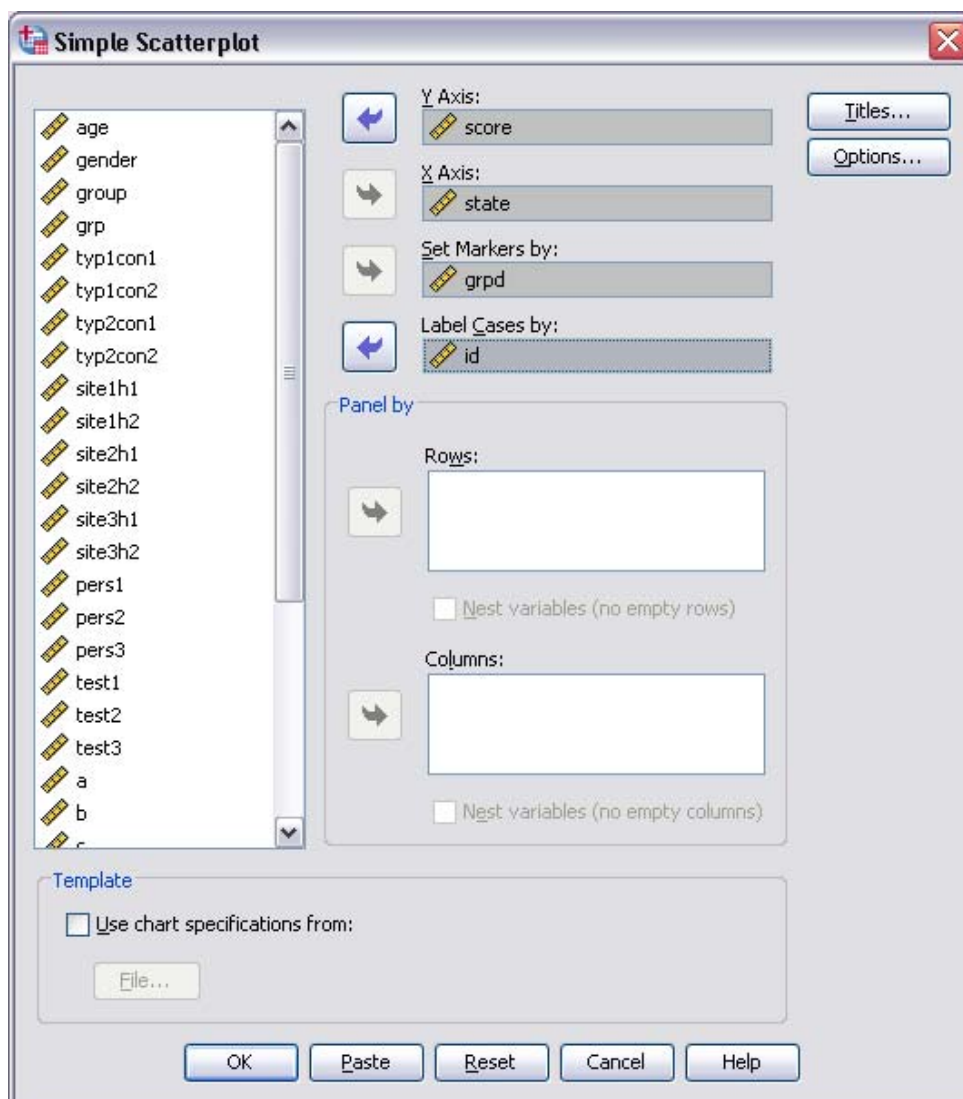
Interactions are most easily understood if they are shown in a graph. It is also a good idea to check interactions by looking at an appropriate scatterplot – you may find that what looks like quite a convincing effect in terms of its statistical significance, or even its effect size, is actually due to a few stray points, especially in small datasets.

Unfortunately, you can't use the plotting facilities provided by *GLM* if one of your independent variables is continuous rather than categorical (i.e., is specified as a covariate rather than a fixed factor)

##### (i) When the Variables are Not Adjusted for Other Variables


This part of the section goes through the steps involved in producing a scatterplot like the one shown earlier in this section. This method applies only if there are no other variables in the model, and you therefore don't need to plot values adjusted for other variables.

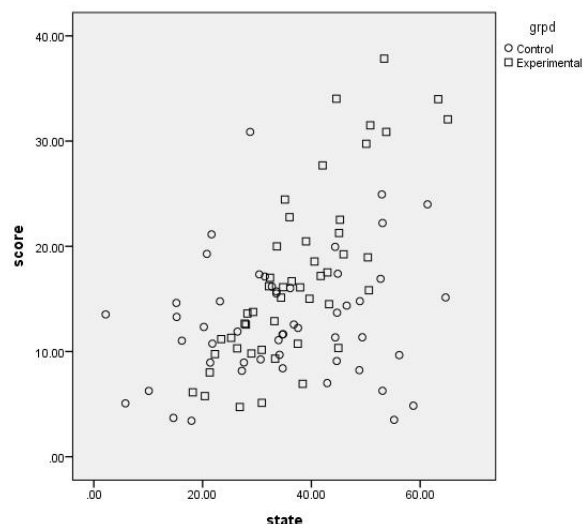
Click on *Graphs* → *Legacy Dialogs* → *Scatter/Dot* → *Simple Scatter* → *Define*. The following display will appear.

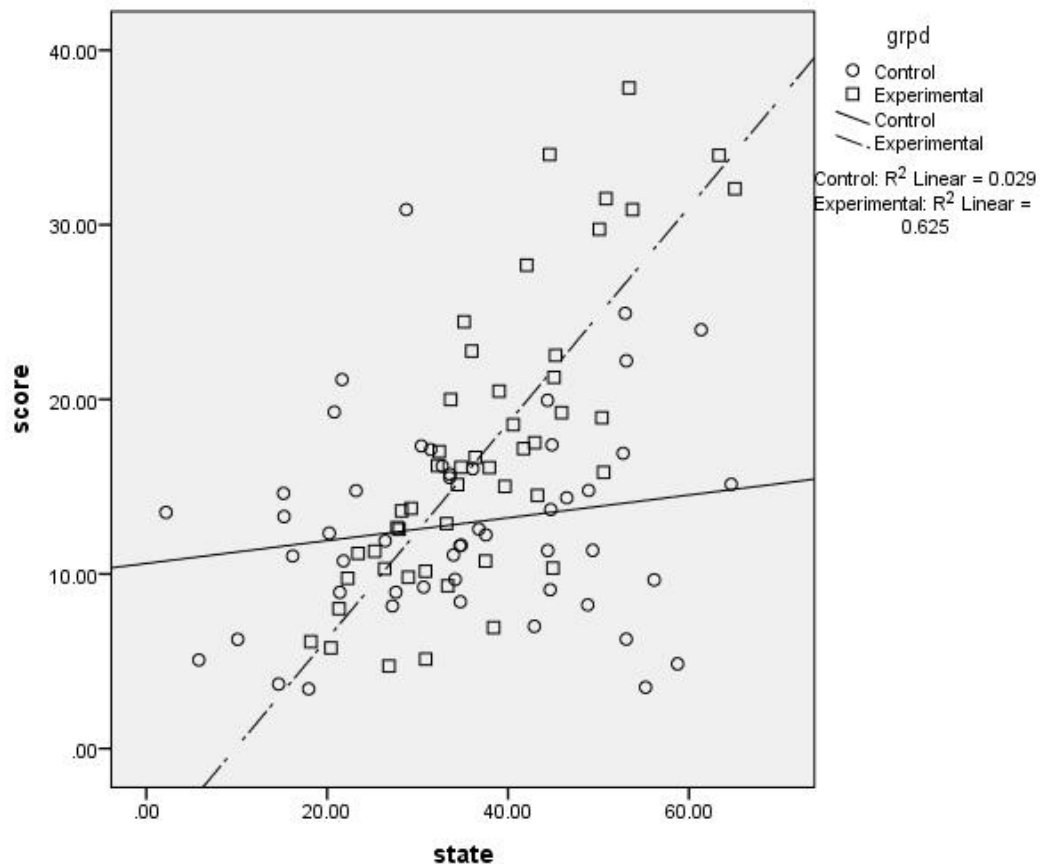


Select *score* for the *Y Axis* and *state* for the *X Axis*. Select *grp* for *Set Markers by*, which means that different symbols will be used for subjects from different groups. Select *id* for *Label Cases by* so that you can easily identify any aberrant or interesting cases which appear in the plot. Click on *OK*.

The graph appears as follows

Double-click on the image to open it in the Chart Editor. Click on the *Add Fit Line at Subgroups* icon  and separate regression lines will be added for each group, as shown on the next page.





(ii) When the Variables are Adjusted for Other Variables

If there had been variables in the model other than those involved in the interaction (*grp**d* and *state*), a graph like the one described in the previous section could be misleading, because no adjustment is made for the other variables, and the form of the interaction could be different from what is shown in the graph. When there are other variables, there's nothing for it but to use the regression equation to obtain predicted values, which can be plotted.

There are two ways of doing this, the hard way and the easy way. The hard way will be illustrated first, for heuristic purposes, then we'll breeze through the easy way, which is provided by *GLM*.

### The Hard Way

For our purposes, we'll run the previous analysis, but with the addition of two variables, *gend* and *age*. To do this analysis, set up *glm* as before, but add *gend* and *age* as further covariates. We'll also centre *age* (behind the scenes I've found that the mean age is 24.76). The reason for this will become obvious. Make sure that the interaction between *grp**d* and *state* is specified in the *Model* section. The syntax for the analysis is:

```
temporary.
compute age=age - 24.76.
glm score with gend age grp d state/
```



```
print=parameters/
design=gend age grpd state grpd*state.
```

The parameter estimates are shown in the table below. The part of the regression equation we're interested in is

$$\text{score} = 13.672 - .10.86 * \text{grpd} - .0049 * \text{state} + .3329 * \text{grpd} * \text{state}.$$

For the purposes of evaluating the equation, we're going to assume that *gend* is equal to zero (female), so we don't need to include it in the equation. (As an aside, an alternative would be to use .576, the proportion of males in the sample as the value for *gend*; this would have the effect of producing predicted values for a population with the same proportion of males and females as the sample. For that, matter we could also substitute .5, which would provide values for the general population.) Also, because *age* was centred, so that the mean age is now zero, that can be dropped from the equation too without producing nonsensical predictions of *score* (you come to love the value zero when substituting values in a regression equation).

**Parameter Estimates**

Dependent Variable: SCORE

Parameter	B	Std. Error	t	Sig.
Intercept	13.6720	1.725	7.928	.000
GEND	1.1008	.854	1.288	.201
AGE	.5832	.069	8.471	.000
GRPD	-10.8593	2.702	-4.019	.000
STATE	-.0049	.041	-.121	.904
GRPD * STATE	.3329	.072	4.595	.000

In order to produce a graph with two straight lines, two points are needed to define each line. The values substituted in the equation for *grpd* will be zero and 1. What values should be used for *state*? Probably values +/- one standard deviation from the mean will provide sensible estimates. Earlier in this section we obtained a table which gave the mean and standard deviation of *state* as approximately 36 and 13 respectively. Plus and minus one standard deviation from the mean gives 23 and 49 respectively. Inspection of the scatterplot shows that these values are well within the range of the observed values – if they weren't, +/- half a standard deviation could be used.

The equation, and the values which are to substituted into it, are shown in the table below.

Term	Intercept	<i>grpd</i>	<i>state</i>	<i>grpd * state</i>	Predicted value
Coeff.	13.67	-10.86	-.0049	.3329	
Values	1	0	23	0 * 23	13.56
	1	0	49	0 * 49	13.43
	1	1	23	1 * 23	10.35
	1	1	49	1 * 49	18.88

The predicted values, which can be plotted, are shown in the final column of the table. They are obtained by the following type of calculation:

Predicted value of *score* =  $13.67 - 10.86 * 0 - .0049 * 23 + .3329 * 0 * 23 = 13.56$

### The Easy Way

As discussed in section 9.3.2, it is possible to produce marginal means for groups, or levels of factors, and to test the significance of the differences between them, at specified values of covariates. We can use this facility to provide the means needed for interaction plots. The following commands give rise to the same means as those produced manually above:

```
glm score by grpd with gend age state/
print=parameters/
emmeans=table(grpd) compare(grpd) with(gend=0 age=24.76 state=23)/
emmeans=table(grpd) compare(grpd) with(gend=0 age=24.76 state=49)/
design=gend age grpd state*state.
```

The relevant tables are:

#### Estimates

Dependent Variable:score

grpd	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0 Control	13.559 <sup>a</sup>	.979	11.615	15.504
1 Experimental	10.356 <sup>a</sup>	1.207	7.959	12.754

a. Covariates appearing in the model are evaluated at the following values: gend = 0, age = 24.76, state = 23.00.

#### Estimates

Dependent Variable:score

grpd	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
0 Control	13.432 <sup>a</sup>	.878	11.688	15.177
1 Experimental	18.884 <sup>a</sup>	1.117	16.665	21.103

a. Covariates appearing in the model are evaluated at the following values: gend = 0, age = 24.76, state = 49.00.

As can be seen, the means shown in the table are identical (within rounding error) to those produced manually.

### Plotting the Graph

The next step is to plot the graph. If you don't have a favourite graphing program you'd rather use, try this method in SPSS.

- (i) Click on *File* → *New* to create a clear data window.

(ii) Click on the *Variable View* tab.

(iii) Enter three variable names in the *Name* column: *grp*, *state* and *score*.

(iv) In the *Decimals* column, enter zero for *grp* and *state*.

(v) In the *Values* column, enter 0 'Control' and 1 'Experimental' for *grp*, and -1 'Mean – 1 SD' and 1 'Mean + 1 SD'.

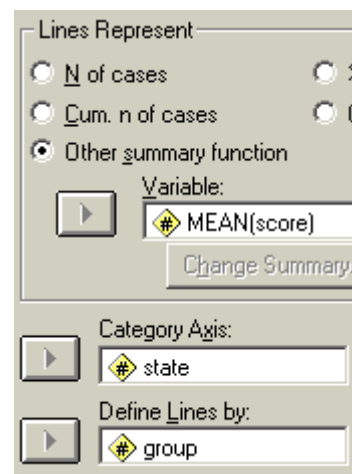
(vi) Click on the *Data View* tab.

(vii) Enter the values as shown:

	group	state	score
1	0	-1	13.56
2	0	1	13.43
3	1	-1	10.36
4	1	1	18.88
5			
6			

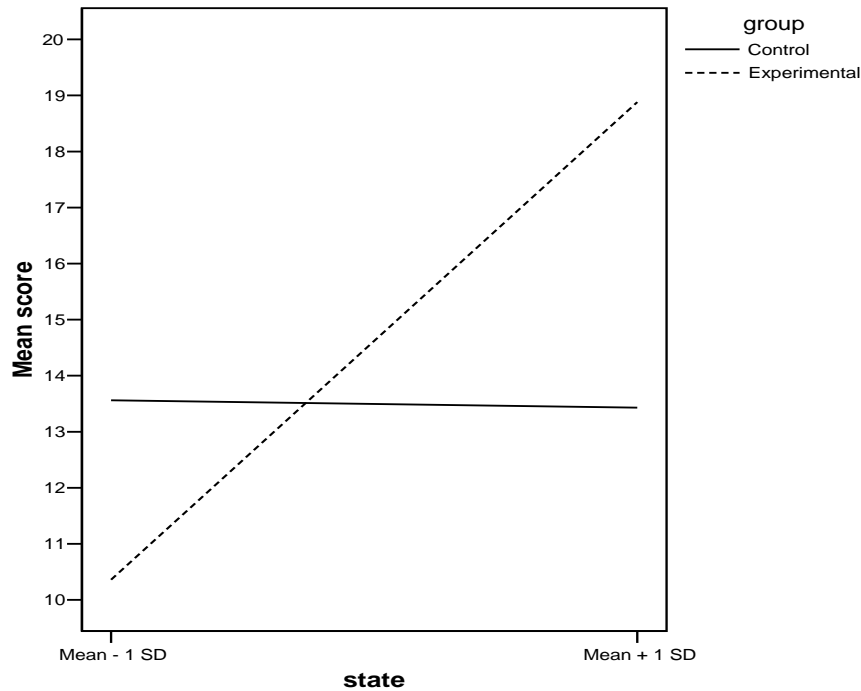
(ix) Click on *Graphs* → *Legacy Dialogs* → *Line*, and select *Multiple* and *Summaries for groups of cases* (the default). Click on *Define*.

(x) Select the variables as shown:



(xi) Click on *OK*.

(xii) Contemplate the graph:



(xiii) Edit the lines, labels, colours etc as you wish by double-clicking on the graph and using the facilities of the SPSS Graphic Editor.

## 15. The Interaction Between Two Numeric Variables

When you're interested in the interaction between two numeric variables, try to resist the temptation to categorise one or both of them, unless you have a good reason to do so. In particular, be aware that if you find an interaction between two numeric variables, then categorise the two variables for further investigation, you're actually investigating a different effect from the one you obtained in the first place. For a thorough study of the evils of dichotomising numeric variables in particular, see MacCallum *et al* (2002).

This example shows how to investigate an interaction between two numeric variables, *age* and *state2*. The dependent variable is *tc*. It is hypothesised that the *tc* score for a given subject will depend on the age of the subject and his/her score on a measure of trait anxiety. It is further hypothesised that the effect of the *state2* measure will vary with age, i.e., that *age* and *state2* will interact. Gender will also be included in the analysis. The means and standard deviations of *age* and *state2* are shown below.

*descriptives age state2.*

	N	Minimum	Maximum	Mean	Std. Deviation
AGE	99	15.00	50.00	24.7576	7.59040
STATE2	99	17.96	84.96	31.0672	11.73095

Point-and-click: *Analyze* → *General Linear Model* → *Univariate*. Specify *tc* as the *Dependent Variable*, *gend*, *age* and *state2* as *Covariates*, and check *Parameter estimates* as an *Option*. Click on the *Model* button, then click on the *Custom* radio button. Select

*gender*, *age* and *state2*, select *Main Effects* in the *Build Terms* slot, and click on the arrow. Repeat for *age* and *state2*, but this time select *Interaction* in the *Build Terms* slot.

#### Syntax

```
glm tc with gend age state2/
print=parameters/
design=gend age state2 age*state2.
```

The results are as follows:

#### Parameter Estimates

Dependent Variable: TC

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	3.1032	.438	7.082	.000	2.2332	3.9732
GEND	-.0737	.082	-.899	.371	-.2366	.0891
AGE	-.1087	.016	-6.796	.000	-.1404	-.0769
STATE2	.1771	.016	11.115	.000	.1455	.2087
AGE * STATE2	-.0015	.000	-3.738	.000	-.0022	-.0007

There is an interaction between *age* and *state2*. However, before we go any further, wouldn't it be a good idea to centre both of these variables, so that the coefficients for *age* and *state2* show their effects at a sensible value of the other variable? For example, the coefficient of .1771 for *state2* in the above table shows its effect for an age of zero. Likewise, it can be seen from the above table of descriptive statistics that zero is well outside the range of values observed for *state2*, so the coefficient for *age* is also unlikely to be sensible. Repeat the analysis with the following syntax or its point-and-click equivalent:

*temporary.*

*compute age=age - 24.76.*

*compute state2=state2 - 31.07.*

*glm tc with gend age state2/*

*print=parameters/*

*design=gend age state2 age\*state2.*

The new table is:

#### Parameter Estimates

Dependent Variable: TC

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4.7945	.065	73.591	.000	4.6651	4.9239
GEND	-.0737	.082	-.899	.371	-.2366	.0891
AGE	-.1539	.011	-14.158	.000	-.1755	-.1323
STATE2	.1410	.008	16.730	.000	.1243	.1578
AGE * STATE2	-.0015	.000	-3.738	.000	-.0022	-.0007

The values in this table say that (i) at the mean of *state2* an increase of a year in age is associated with a drop of .1539 in *tc* (ii) at the mean of *age* an increase of one in *state2* is associated with an increase in *tc* (so much for the idea that higher anxiety will lead to worse performance), and (iii) when age increases by a year, the effect of *state2* on *tc* is reduced by .0015 (likewise, as *state2* increases by one, the effect of *age* on *tc* is decreased by .0015). The interaction effect is therefore pretty subtle compared to the effects of the two individual variables. (The partial  $\eta^2$  values [not shown in the output] for *age*, *state2* and *age\*state2* respectively are .68, .75 and .13.)

A graph would help! In order to represent what is a continuous process in an understandable way, we need to pick out some values of the two variables which can be substituted into the regression equation in the same way as we did for *state* in the previous example. The hypothesis stated at the beginning of the section was that the effects of *state2* on *tc* would be moderated by age. Therefore, in the graph, *state2* should be on the *x*-axis and that there should be separate lines for several values of *age*.

### The Hard Way

Because the analysis was done without and then with centred versions of *age* and *state2*, we have the choice of substituting raw values into the equation for the first analysis or deviations from the mean into the equation for the second analysis. Because we used the raw values in the previous section (*state* wasn't centred), we'll use deviations here.

From the above table, the standard deviations of *age* and *state2* are approximately 8 and 12 respectively. (As an aside, it can be seen from the minimums and maximums shown in the table above that the distributions of both *age* and *state2* are positively skewed, so that one standard deviation below the mean is close to the minimum for both variables. If they were actually outside the range of the observed values, we'd think of using +/- half a standard deviation.)

Because *state2* is to be on the *x*-axis, we only need to substitute two values in the equation; we'll use -12 and 12. For *age*, as the moderating variable, we'd like a few more values, so we'll use +/- both one, and a half, standard deviation, plus the mean itself (-8, -4, 0, 4 and 8). Because there are two values for *state2* and five for *age*, there will be 10 estimates of *tc*. *Gend* will be set to zero and omitted from the equation, so that the estimates produced will be for females although, because the effect of gender is small, this won't make much difference. The calculations are shown in the table:

Term	Intercept	<i>age</i>	<i>state2</i>	<i>age * state2</i>	Predicted value
Coeff.	4.7945	-.1539	.141	-.0015	
Values	1	-8	-12	-8 * -12	4.19
	1	-8	12	-8 * 12	7.86
	1	-4	-12	-4 * -12	3.65
	1	-4	12	-4 * 12	7.17
	1	0	-12	0 * -12	3.10
	1	0	12	0 * 12	6.49
	1	4	-12	4 * -12	2.56
	1	4	12	4 * 12	5.80

	1	8	-12	8 * -12	2.02
	1	8	12	8 * 12	5.11

### The Easy Way

We can again use the estimated marginal means subcommand to obtain the values to plot. With this method, there is no need to centre *age* and *state2*; instead, we simply specify the actual values of the variables which we want to substitute into the regression equation (the mean and +/- half and +/- one standard deviation for *age*, and +/- one standard deviation from the mean for *state2*):

```
glm tc with gend age state2/
  print=parameters/
  emmeans=table(overall) with(gend=0 age=16.76 state2=19.0676)/
  emmeans=table(overall) with(gend=0 age=20.76 state2=19.0676)/
  emmeans=table(overall) with(gend=0 age=24.76 state2=19.0676)/
  emmeans=table(overall) with(gend=0 age=28.76 state2=19.0676)/
  emmeans=table(overall) with(gend=0 age=32.76 state2=19.0676)/
  emmeans=table(overall) with(gend=0 age=16.76 state2=43.0676)/
  emmeans=table(overall) with(gend=0 age=20.76 state2=43.0676)/
  emmeans=table(overall) with(gend=0 age=24.76 state2=43.0676)/
  emmeans=table(overall) with(gend=0 age=28.76 state2=43.0676)/
  emmeans=table(overall) with(gend=0 age=32.76 state2=43.0676)/
  design=gend age state2 age*state2.
```

Note that, because we haven't specified any factors, the *table* keyword doesn't contain the name of a factor, but the term *overall*, which is simply the predicted value of *tc*, given the specified values of *gend*, *age* and *state2*.

The series of *emmeans* subcommands produces a set of single tables, each showing the estimated mean corresponding to the specified values of the covariates. For example:

#### 1. Grand Mean

Dependent Variable:tc

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
4.193 <sup>a</sup>	.102	3.991	4.395

a. Covariates appearing in the model are evaluated at the following values: gend = 0, age = 16.76, state2 = 19.07.

#### 6. Grand Mean

Dependent Variable:tc

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
7.858 <sup>a</sup>	.213	7.435	8.280

a. Covariates appearing in the model are evaluated at the following values: gend = 0, age = 16.76, state2 = 43.07.

#### 3. Grand Mean

Dependent Variable:tc

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
3.102 <sup>a</sup>	.110	2.884	3.319

a. Covariates appearing in the model are evaluated at the following values: gend = 0, age = 24.76, state2 = 19.07.

#### 8. Grand Mean

Dependent Variable:tc

Mean	Std. Error	95% Confidence Interval	
		Lower Bound	Upper Bound
6.487 <sup>a</sup>	.130	6.228	6.745

a. Covariates appearing in the model are evaluated at the following values: gend = 0, age = 24.76, state2 = 43.07.

These means, and the others which are not shown, are the same as those we produced manually, within rounding error.

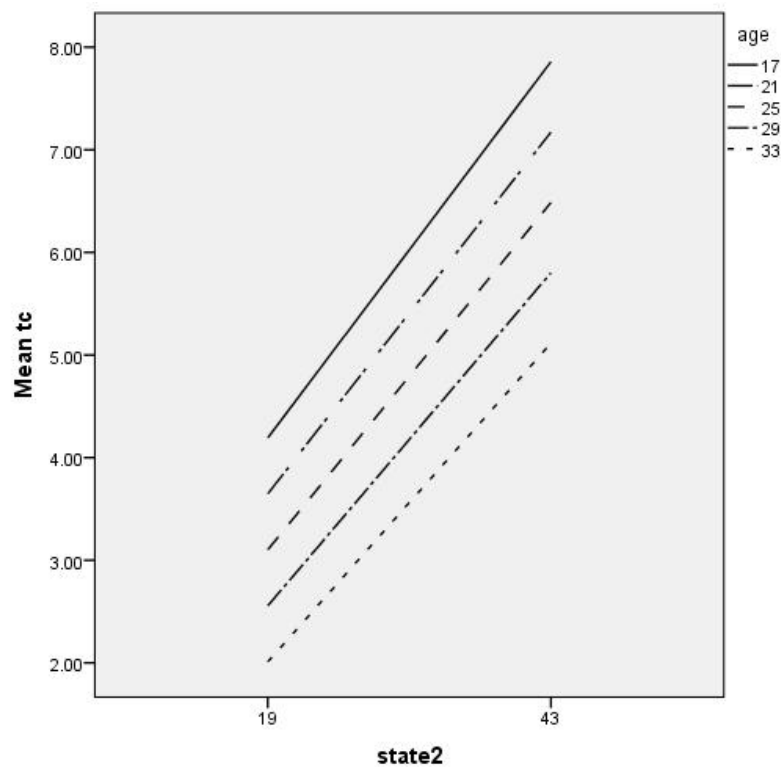
### Plotting the Graph

The data can be entered into SPSS as in the previous example:

	age	state2	tc
1	17	19	4.19
2	17	43	7.86
3	21	19	3.65
4	21	43	7.17
5	25	19	3.10
6	25	43	6.49
7	29	19	2.56
8	29	43	5.80
9	33	19	2.01
10	33	43	5.12

Notice that the actual values of *age* and *state2* have been entered, rather than the deviations. This may make the graph more easily understood.

The graph is shown below:





The interaction effect is certainly subtle! This is partly because, as mentioned earlier, it's small in comparison to the effects of *age* and *state2* (note the clear differences in terms of *tc* between the *age* lines, and the slope of the lines representing the effect of *state2*).

It's also true that the graph doesn't represent the higher values of *age* and *state2*; if it did, the differences in the slope of the lines would be more obvious.

Something like the following could be said about this result: 'Younger subjects, and those with higher scores on the state variable, received higher *tc* scores. There was also a slight tendency for the effect of *state2* scores on *tc* to be less marked for older subjects than for young subjects. This interaction effect was statistically significant ( $t(94) = 3.74, p < .0005$ , but was not of great practical significance compared to the main effects of *age* and *state2*'.

### **Box 9: 'Pick a point' and Regions of Significance**

The preceding section has described a method for testing the simple slope of variables involved in interactions at specific values of the other variable involved in the interaction. This approach has been dubbed the 'pick a point' method by Bauer & Curran (2005), a term which highlights the arbitrary nature of the procedure when applied to numeric as opposed to categorical variables. For this reason, the method is more suited for illustrative than hypothesis-testing purposes (unless the researcher happens to have hypotheses about specific values).

Bauer and Curran discuss the use of the Johnson-Neyman 'regions of significance' procedure, which was described in section 3.3, for inferential purposes, and generalise it for use in various contexts. Together with the very nice SPSS macro provided by Hayes & Matthes (2009), this work makes the J-N method more accessible, and it will no doubt be used more often than it has been.



## References

- Bauer, D.J., & Curran, P.J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, 40, 373-400.
- Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Harris, R.J. (1994). *ANOVA: An analysis of variance primer*. Itasca, Illinois: Peacock.
- Hayes, A. F., & Matthes, J. (2009). Computational methods for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods*, 41, 924-936.
- Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association*, 81, 1000-1004.
- Huynh, H., & Feldt, L. (1976). Estimation of the Box corrections for degrees of freedom from sample data in randomized block and-split plot design. *Journal of Educational Statistics*, 1, 69-82.
- Johnson, P.O., & Fay, L.C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, 15, 349-367.
- Johnson, P.O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1, 57-93.
- MacCallum, R.C.et.al. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, (1), 19-40.
- Mauchly, J. (1940). Significance test for sphericity of a normal  $n$ -variate distribution. *Annals of Mathematical Statistics*, 11, 204-209.
- Maxwell, S.E. & Delaney, H.D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, California: Wadsworth.
- Ramsey, P. H. (2002). Comparison of closed testing procedures for pairwise testing of means. *Psychological Methods*, 7, 504-523.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, N.J.: L. Erlbaum Associates.
- Winer, B., Brown, D., & Michels, K. (1991). *Statistical principles in experimental design* Third Edition. New York, McGraw-Hill.



## Appendix 1: Variables in *glmdemo.sav*.

Up to *test3*, the variables are based on an actual dataset, but variables were renamed, random noise was added, and differences were created, using commands like the following:

```
do repeat i=typ1con1 to pers6.
compute i=i + rv.normal(0,.25).
end repeat.
```

```
do repeat i=typ1con1 to pers6.
compute i=i + group * abs(rv.normal(0,.25)).
compute i=i + grp * abs(rv.normal(0,.25)).
end repeat.
```

```
compute age=rnd(age - (group - 5) * abs(rv.normal(0,.25))).
```

Variables after *test3* are either dummy-coded versions of the original variables (e.g., *grp*), averages over two or more of the original variables (e.g., *tc*), completely random (e.g., *a*, *b* and *c*), random but with properties based on other variables (e.g., *state*, *score*) or recodes of generated variables (e.g., *brief*).

Variable	Order	Value labels	Comments
ID	1		
AGE	2		
GENDER	3	1 male 2 female	
GROUP	4	1 Control 2 25 mg/kg 3 50 mg/kg 4 100 mg/kg	
GRP	5	1 Control 2 Experimental	
TYP1CON1	6		Crossed within-subject factors <i>type</i> (2 levels) and <i>condition</i> (2).
TYP1CON2	7		
TYP2CON1	8		
TYP2CON2	9		
SITE1H1	10		Crossed within-subject factors <i>site</i> (3) and <i>hemisphere</i> (2).
SITE1H2	11		
SITE2H1	12		
SITE2H2	13		
SITE3H1	14		
SITE3H2	15		
PERS1	16		Three measures on personality inventories.
PERS2	17		
PERS3	18		
TEST1	19		Three test scores.

TEST2	20		Random variables produced with these commands:  <i>compute x=rv.normal(10,1.5).</i> <i>compute a=x + rv.normal(10,1.5).</i> <i>compute b=x + rv.normal(10,1.5) + .25.</i> <i>compute c=x + rv.normal(10,1.5) + .5.</i>
TEST3	21		
A	23		
B	24		
C	25		
SITE1	30		site averaged over <i>hemi</i> by the following commands:  <i>compute site1=mean(site1h1,site1h2).</i> <i>compute site2=mean(site2h1,site2h2).</i> <i>compute site3=mean(site3h1,site3h2).</i>
SITE2	31		
SITE3	32		
GRPD	33	0 Control 1 Experimental	Dummy-coding of <i>grp</i> :  <i>recode grp (1=0)(2=1) into grpd.</i> <i>print format grpd (f1).</i> <i>value labels grpd 0 'Control' 1 'Experimental'.</i>
GEND	34	0 Female 1 Male	Dummy-coding of <i>gender</i> :  <i>recode gender (1=1)(2=0) into gend.</i> <i>print format gend (f1).</i> <i>value labels gend 0 'Female' 1 'Male'.</i>
GRP2	35		Dummy-coding of <i>group</i> :  <i>recode group</i> <i>(2=1)(missing=sysmis)(else=0)</i> <i>into grp2.</i> <i>recode group</i> <i>(3=1)(missing=sysmis)(else=0)</i> <i>into grp3.</i> <i>recode group</i> <i>(4=1)(missing=sysmis)(else=0)</i> <i>into grp4.</i>
GRP3	36		
GRP4	37		
SCORE	38		Created with:  <i>compute score = (-3 - .1 * pers1 - .25 * age + .2 * pers1 * age) + rv.normal(0,3).</i>
STATE	40		<i>do if (gend eq 0).</i> <i>compute state=rv.normal(50,20).</i> <i>else if (gend eq 1).</i>

			$compute\ state = (2 * score + 20) +$ $rv.normal(0,6).$ $end\ if.$
BRIEF	41	0 No briefing 1 Briefing	Automatic recode of <i>state</i> :  $RANK$ $VARIABLES = state$ $/NTILES(2)$ $/PRINT = NO$ $/TIES = MEAN.$
TC	42		Mean of <i>typ1con1</i> to <i>typ2con2</i> . Created with:  $compute\ tc = mean(typ1con1\ to\ typ2con2).$
STATE2	43		Created with:  $compute\ state2 = (3.75 -.257 * tc -.005 * age + .25 * tc * age) + rv.normal(0,2).$





## Appendix 2: Calculation of $\eta^2$ with Three IVs

In this example, we're interested in calculating  $\eta^2$  for *pers2* for a main effects model which also contains *gender* and *grp* as IVs. *Test2* is the DV.

In the first analysis, Type III (unique) Sums of Squares (SS) are used. The SS for *pers2* is 3.542466. We want to calculate  $\eta^2$  for *pers2* while adjusting for the other two

### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.565600 <sup>a</sup>	3	2.522	4.297	.007	.119484
Intercept	34.559807	1	34.560	58.888	.000	.382667
GENDER	2.074889	1	2.075	3.535	.063	.035880
PERS2	3.542466	1	3.542	6.036	.016	.059742
GRP	1.427569	1	1.428	2.432	.122	.024966
Error	55.753316	95	.587			
Total	2410.015407	99				
Corrected Total	63.318916	98				

a. R Squared = .119 (Adjusted R Squared = .092)

variables. We can't use the SSs for *gender* and *grp* given in the first ANOVA table, because they're adjusted for each other and for *pers2*. So we do the analysis shown in the

### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.565600 <sup>a</sup>	3	2.522	4.297	.007	.119484
Intercept	2346.6965	1	2346.696	3998.617	.000	.976793
GENDER	2.175082	1	2.175	3.706	.057	.037548
GRP	1.848052	1	1.848	3.149	.079	.032083
PERS2	3.542466	1	3.542	6.036	.016	.059742
Error	55.753316	95	.587			
Total	2410.0154	99				
Corrected Total	63.318916	98				

a. R Squared = .119 (Adjusted R Squared = .092)

next table, in which *gender* and *grp* are entered before *pers2* in a Type I (sequential) analysis. Note the the SS for *pers2* is the same as in the first analysis, but those for *gender* and *grp* have changed. The calculation of  $\eta^2$  for *pers2* is

$$\begin{aligned}
 \eta^2 &= SS(pers2) / SS(total) - SS(grp + gender) \\
 &= 3.542466 / 63.318916 - (2.175082 + 1.848052) \\
 &= 3.542466 / 59.295782 \\
 &= .059742
 \end{aligned}$$

which is the value given in both the tables. Note that if *gender* and *grp* are entered in a different order, as in the table below, the SS for each variable is different, but the sum of the two SSs is the same, and the value of  $\eta^2$  for *pers2* is consequently the same.

#### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.565600 <sup>a</sup>	3	2.522	4.297	.007	.119484
Intercept	2346.6965	1	2346.696	3998.617	.000	.976793
GRP	1.788502	1	1.789	3.047	.084	.031082
GENDER	2.234632	1	2.235	3.808	.054	.038536
PERS2	3.542466	1	3.542	6.036	.016	.059742
Error	55.753316	95	.587			
Total	2410.0154	99				
Corrected Total	63.318916	98				

a. R Squared = .119 (Adjusted R Squared = .092)

In the final analysis, shown below, the effect of *pers2* is assessed in a Type I (sequential) analysis. In this example we want to adjust for *gender*, but not for *grp*.  $\eta^2$  can be

#### Tests of Between-Subjects Effects

Dependent Variable: TEST2 Test score 2

Source	Type I Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.565600 <sup>a</sup>	3	2.522	4.297	.007	.119484
Intercept	2346.6965	1	2346.696	3998.617	.000	.976793
GENDER	2.175082	1	2.175	3.706	.057	.037548
PERS2	3.962950	1	3.963	6.753	.011	.066363
GRP	1.427569	1	1.428	2.432	.122	.024966
Error	55.753316	95	.587			
Total	2410.0154	99				
Corrected Total	63.318916	98				

a. R Squared = .119 (Adjusted R Squared = .092)

calculated directly from the output in the table:

$$\begin{aligned}
 \eta^2 &= SS(pers2) / SS(total) - SS(grp + gender) \\
 &= 3.962950 / 63.318916 - (2.175082 + 1.427569) \\
 &= 3.962950 / 63.318916 - (2.175082 + 1.427569) \\
 &= 3.962950 / 59.716265 \\
 &= .066363
 \end{aligned}$$

which is the value shown in the table. Note that although the SS for *pers2* are adjusted only for *gender*, the SS for *grp* enters the denominator in the calculation of  $\eta^2$  but the SS used are adjusted for both *gender* and *pers2*.

The general rule is that the SS used for the denominator are for all the variables in the model except the variable of interest, but the SS used are appropriate for the Type of SS used in the original analysis. In the first analysis, *gender* and *grp* had to be fitted first, but in order to get the appropriate SS (i.e., not adjusted for *pers2* and not simultaneously adjusted for each other), we had to perform a Type I analysis with *gender* and *grp* fitted before *pers2*.

In the second, sequential analysis (fourth table above), the SS for *gender* and *grp* were those obtained for the variables fitted in the order in which they occurred.



### Appendix 3: The Matrix Multiplication in Section 6.2.1.5

Let Matrix 1 be the matrix of the orthonormalised polynomial contrast coefficients:

-.7071	0	.7071
.4082	-.8165	.4082

Let Matrix 2 be the variance-covariance matrix for the variables  $a$ ,  $b$  and  $c$ :

	$a$	$b$	$c$
$a$	4.7094	2.3975	2.3793
$b$	2.3975	4.6267	2.3034
$c$	2.3793	2.3034	4.6567

Let Matrix 3 be the transpose of Matrix 1 (i.e., Matrix 1 is 'flipped' so rows become columns and columns become rows):

-.7071	.4082
0	-.8165
.7071	.4082

The crucial property is that if Matrix 2 (the variance-covariance matrix of the variables defining the within-subject factor) is circular, then

Matrix1 \* Matrix 2 \* Matrix3 = Matrix4 has the property of sphericity.

Multiplying Matrix1 by Matrix2

The numbers in each row of Matrix1 are multiplied by the numbers in each column of Matrix2 and summed:

$$-.7071 * 4.7094 + 0 * 2.3975 + .7071 * 2.3793 = -1.6476$$

$$-.7071 * 2.3975 + 0 * 4.6267 + .7071 * 2.3034 = -0.0665$$

$$-.7071 * 2.3793 + 0 * 2.3034 + .7071 * 4.6567 = 1.6104$$

$$.4082 * 4.7094 - .8165 * 2.3975 + .4082 * 2.3793 = 0.9360$$

$$.4082 * 2.3975 - .8165 * 4.6267 + .4082 * 2.3034 = -1.8588$$

$$.4082 * 2.3793 - .8165 * 2.3034 + .4082 * 4.6567 = 0.9914$$

The resulting matrix is

-1.6476	-0.0665	1.6104
0.9360	-1.8588	0.9914

Multiplying the Result of (Matrix1 \* Matrix2) by Matrix3 (which is Matrix 1 transposed)

$$-1.6476 * -.7071 - .0665 * 0 + 1.6104 * .7071 = 2.3037$$

$$-1.6476 * .4082 - .0665 * -.8165 + 1.6104 * .4082 = 0.0391$$

$$.9360 * -.7071 - 1.8588 * 0 + .9914 * .7071 = 0.0392$$

$$.9360 * .4082 - 1.8588 * -.8165 + .9914 * .4082 = 2.3045$$

The resulting matrix (Matrix 4):

2.3037	0.0391
0.0392	2.3045