Intermediate Analysis of Variance with SPSS

James M. Clark

TABLE OF CONTENTS

PREFACE

This is the second of two manuscripts. The first reviewed basic statistics and then expanded on regression, specifically introducing students to multiple regression and its applications. The current manuscript assumes a basic understanding of statistics. The correspondence between ANOVA and multiple regression is demonstrated in places, but constitute distinct sections that can be omitted if students are not yet familiar with multiple regression. However, certain aspects of ANOVA (e.g., contrast analyses) do benefit from an understanding of regression.

And slightly edited from the Preface of the regression manuscript ...

ANOVA as presented in this manuscript involves only basic mathematical operations. Understanding material like statistics requires practice and repetition, and also benefits from exposure to alternative conceptualization of important features, such as interaction effects. Although, conceptual repetition can seem confusing and redundant, it results in a deeper understanding of the analyses.

With respect to SPSS, I focus on syntax, which is to be recommended over a menu approach. It provides a record of the analyses, makes it easy to correct and rerun analyses, allows creation of simulations to generate data, and gives access to some procedures not available by menu (e.g., MANOVA).

Thanks to several colleagues who have contributed to my own understanding of regression and to the many students over the years who tolerated my sometimes "casual" lectures on this material. Errors or suggestions? Please e-mail j.clark@uwinnipeg.ca. Thanks ... Jim

CHAPTER 1 - SINGLE FACTOR BETWEEN-SUBJECTS OMNIBUS F TEST

Analysis of Variance (ANOVA) is typically used when research designs have the following characteristics: relatively few predictors (i.e., factors, independent variables), each factor involves a small number of levels or conditions, multiple factors are uncorrelated with one another (i.e., independent, orthogonal), and levels of the factors are unordered (i.e., categorical), although ordered (i.e., numerical) factors can be accommodated with appropriate supplementary analyses. Factors can be experimental manipulations (e.g., Treatment, Control) or natural, pre-existing differences (e.g., Young, Middle Age, Old). An equal number of subjects in each condition is also common and can simplify analyses.

ANOVA designs include one or more of two types of factors. ***Between-Subject*** factors compare groups that involve independent or uncorrelated observations; that is, there is no correlation between scores at different levels of the factor. Between-S factors could be different experimental conditions for unrelated subjects (e.g., different instructions in a memory task for two or more groups of subjects, different treatments for psychopathology applied to different clients, different methods of teaching reading used in different classrooms, ...), or pre-existing individual differences such as gender (two levels) if they involve uncorrelated observations. ANOVA for a single Between-S factor with two levels (e.g., Treatment vs. Control; Male vs. Female) is equivalent to an independent groups t-test. But ANOVA is required for more than two levels (e.g., Control vs. Placebo vs. CBT; Young vs. Middle Age vs. Old with unrelated participants at each age level).

The second type of factor in ANOVA designs are ***Within-Subject*** factors, which involve observations that are correlated at different levels of the factors. For example, each subject in an experiment could be exposed to all conditions (e.g., Pretest vs. Posttest designs; memory for lists that contain both Concrete vs. Abstract words). Or within-S factors can compare pre-existing groups (e.g., Male vs. Female; Young vs. Middle Age vs. Old) if scores are correlated (e.g., twins, longitudinal study of the same people at different ages). ANOVA for a single Within-S factor with two levels is equivalent to a paired-difference t-test. ANOVA is required for more than two levels (e.g., Pre vs. Post vs. Follow-up).

From a statistical perspective, ANOVA designs vary in the number of factors and whether each factor is Between-S or Within-S. Studies with two or more factors are called factorial designs. A comparison of recall given subjects assigned randomly to one of four instructional conditions is a single-factor Between-S design. A study comparing pretest, posttest, and follow-up depression scores for people randomly assigned to a control group or a treatment group is a two-factor mixed design with one Between-S (Control vs. Treatment) and one Within-S (Pre vs. Post vs. Follow-up) factor.

Ideally in a factorial design, the multiple factors are orthogonal to one another (i.e., independent, uncorrelated), which requires that all levels of each factor occur with all levels of other factors. For example, a study with two factors each with two levels (e.g., A1 & A2, B1 & B2), would have four conditions or cells defined by all possible combinations of the two factors: A1+B1, A1+B2, A2+B1, and A2+B2. A and B are uncorrelated or orthogonal because every level of A occurs with every level of B. Even with such a design, however, orthogonality can be upset by unequal numbers of observations per cell. Independence is a desirable characteristic of factorial designs, and differentiates such designs from correlated predictors that require multiple regression to determine unique effects. Numerous possible ANOVA designs exist. The first three chapters examine the most basic design, a single-factor Between-S design that involves two or more groups and observations that are independent across groups.

Robert Hare, developer of the Psychopathy Check List (PCL), wrote a book called "Snakes in Suits" summarizing evidence that some business leaders demonstrate characteristics of psychopaths related to their success. To examine this hypothesis, specifically whether business programs develop such traits or they pre-exist in students who choose business, organizational psychologists administered a measure of psychopathy to 20 university students, five from each of four majors: humanities (group 1), social science (2), natural science (3), and business (4). Students in the four majors were unrelated, which means that observations were uncorrelated or independent across groups. The results appear below, along with descriptive statistics for each group indicated by a subscript *j* = 1, 2, 3, or 4. In general, *j* = 1, 2, ..., k, where k is the number of levels of the factor. Descriptive statistics are also given for the overall sample, indicated by a subscript G for grand (e.g., $\overline{y}_G$ is the grand mean).

		Major (j = 1, 2, 3, 4)			
j	**1**	**2**	**3**	**4**	
	2	1	5	9	
	6	5	5	8	
	4	3	8	7	
	1	1	2	7	
	7	5	5	9	
$\overline{y}_j$	$\overline{y}_1$=4.0	$\overline{y}_2$=3.0	$\overline{y}_3$=5.0	$\overline{y}_4$=8.0	$\overline{y}_G$ = 5.0
s_j	s_1=2.54951	s_2=2.00000	s_3=2.12132	s_4=1.00000	s_G = 2.65568
n_j	n_1=5	n_2=5	n_3=5	n_4=5	N = 20

Because there is only one factor (i.e., Major) and observations in each group are independent of one another (i.e., there is no reason to expect scores for different majors to correlate), the appropriate analysis is a single factor Between-S or independent groups ANOVA based on the F test. This analysis requires two variances (or mean squares), one for the numerator that represents variability between means for each group, and one for the denominator that represents random or error variability within groups. Calculation of these variances requires a *SS* and a *df* for each.

Consider first the denominator, SS_{Error} or SS_{Within}. Deviations of scores from the group means represent random variation or noise in that we do not know why one person in a condition scored higher or lower on the dependent variable than other people in that same group. Each observation can be represented as y_{ij}, where j represents a group and i identifies individual observations in each group. In our study, for example, y_{42} is the 4th score in group 2 (i.e., $y_{42} = 1$). In the psychopathy study, subject 1 in group 1 obtained a low score of 2 ($y_{11} = 2$) and subject 2 a higher score of 6 (i.e., $y_{21}=6$). Both are Humanities majors, which does not explain the variability in scores. It must be some unknown influence and is treated as error.

$$SS_{Error} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$
$$= \sum_{j=1}^{k} SS_j = (n_j - 1)s_j^2$$
$$df = N - k$$

Box 1-1. SS_{Error}

There are several ways to compute SS within each group, as shown in Box 1-1. The means for each group are represented by $\bar{y}_j$, which becomes $\bar{y}_1$, $\bar{y}_2$, $\bar{y}_3$, and $\bar{y}_4$. The notation in the first line states: subtract the group mean, $\bar{y}_j$, from each y_{ij}, square the resulting deviation, and sum the squared deviations over the number of observations in each group, n_j, and over the number of levels of the factor, $k = 4$ in the present study. The sum of squared deviations of each observation from its group mean is SS for each group, that is $SS_j = \Sigma(y - \bar{y}_j)^2 = (n_j - 1)s_j^2$. These SS_js are summed again over levels of j to get SS_{Error}. To illustrate for group 1:

$$SS_1 = (2-4)^2 + (6-4)^2 + (4-4)^2 + (1-4)^2 + (7-4)^2 = 26.0$$
or $$SS_1 = (n_1 - 1)s_1^2 = (5 - 1)2.54951^2 = 26.0$$

Standard deviations can be used to calculate SS_2, SS_3, and SS_4 and then the four SSs are summed to produce $SS_{Error.}$

$$SS_2 = 4 \times 2.00000^2 = 16.0$$
$$SS_3 = 4 \times 2.12132^2 = 18.0$$
$$SS_4 = 4 \times 1.00000^2 = 4.0.$$
$$SS_{Within} \text{ or } SS_{Error} = \Sigma SS_j = 26.0 + 16.0 + 18.0 + 4.0 = 64.0.$$

The df for SS_{Error} is obtained by summing the df for the separate SS_js; that is,

$$df_{Error} = (5-1) + (5-1) + (5-1) + (5-1) = 16$$
or $$df_{Error} = N - k = 20 - 4 = 16$$

N equals the total number of observations and k equals the number of levels to our factor. Given SS_{Error} and df_{Error},

$$MS_{Error} = SS_{Error}/df_{Error} = 64.0 / 16.0 = 4.0$$

These calculations are a generalization of the formula for s_p^2 for the independent groups t-test, which was: $(SS_1 + SS_2) / (n_1 + n_2 - 2)$.

Turning to the numerator, variability between groups is obtained by calculating the squared deviation of each treatment mean from $\overline{y}_G$ weighted by the number of observations in each group (see Box 1-2). These calculations are illustrated below.

$$SS_{Treatment} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\overline{y}_j - \overline{y}_G)^2 = \sum_{j=1}^{k} n_j(\overline{y}_j - \overline{y}_G)^2$$
$$df = k - 1$$

Box 1-2. Formula for $SS_{Treatment}$.

Major	$\overline{y}_j$	$\overline{y}_j - \overline{y}_G$
1.	4.0	−1.0
2.	3.0	−2.0
3.	5.0	0.0
4.	8.0	+3.0
$\overline{y}_G$	5.0	

$SS_{Between} = \Sigma n_j(\overline{y}_j - \overline{y}_G)^2 = 5\times-1.0^2 + 5\times-2.0^2 + 5\times0.0^2 + 5\times3.0^2 = 70.0$

or if n_js are equal $= 5\times(-1.0^2 + -2.0^2 + 0.0^2 + 3.0^2) = 70.0$

The critical feature of $SS_{Between}$ is that it reflects how much the group means differ from one another. If they were all the same then they would all equal the grand mean ($\overline{y}_G$) and $SS_{Between}$ would be 0. The more the group means vary, the more they deviate from the grand mean, and the larger $SS_{Between}$ becomes. The df for $SS_{Between}$ equals the number of group means minus 1 for the grand mean; i.e., $df_{Between} = k - 1 = 4 - 1 = 3$. Hence,

$MS_{Between} = 70.0 / 3 = 23.333$

The ratio of $MS_{Treatment}$ over MS_{Error} produces an F statistic to test whether there is significant variability in the group means relative to the random variability within groups. As shown in Box 1-3, F tests the null hypothesis that the k population means are all equal. For our study:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_a: one or more equality is false
$$F = \frac{MS_{Treatment}}{MS_{Error}} \quad df = k-1, N-k$$

Box 1-3. Analysis of Variance.

H_0: $\mu_{Humanities} = \mu_{SocialScience} = \mu_{NaturalScience} = \mu_{Business}$
H_a: one or more equality is false
$F_{Observed} = 23.333 / 4.0 = 5.833$ $df = k - 1, N - k = 3, 16$
For $\alpha = .05$, $F_{Critical} = 3.24$ Reject H_0, Accept H_a

Given α = .05, researchers would reject H_0 that the four population means are equivalent, and accept the alternative hypothesis that one or more equality is false. This alternative hypothesis is vague in that it does not specify which groups differ from one another. Follow-up analyses presented in the next chapter allow for more specific conclusions about the nature of the differences among the groups. To distinguish it from more specific comparisons using F, the current F is referred to as the Omnibus F test because it tests the significance of the overall (omnibus) variability in the treatment means.

Before considering how SPSS carries out this analysis, note that ANOVA has partitioned the total variability in scores into variability between groups (the numerator) and variability within groups (the denominator), much as regression partitioned SS_{Total}. That is,

$$SS_{Between} + SS_{Error} = 70.0 + 64.0 = 134.0 = (20-1)2.65568^2 = SS_{Total} = \Sigma\Sigma(y_{ij} - \overline{y}_G)^2$$

Various terms can be used to label the numerator and denominator quantities. The numerator may be referred to as $SS_{Between}$, $SS_{Treatment}$, SS_{Model}, or by the factor name, SS_{Major} in the present case. The denominator may be referred to as SS_{Within} or SS_{Error}, the latter being more general for different designs. These quantities are later shown as well to be equivalent to $SS_{Regression}$ and $SS_{Residual}$.

SPSS and the Single Factor Between-S ANOVA

Like the independent t-test, data for the single factor Between-S design requires two variables per subject, one to indicate which group the observation belongs to and one to indicate the score on the dependent variable. Other variables could be entered (e.g., a subject number) but are not required for ANOVA. Here is the syntax to enter data for the present study.

```
DATA LIST FREE / major psypath.
BEGIN DATA
1 2   1 6   1 4   1 1   1 7         2 1   2 5   2 3   2 1   2 5
3 5   3 5   3 8   3 2   3 5         4 9   4 8   4 7   4 7   4 9
END DATA.
```

Several procedures in SPSS perform ANOVA, including ONEWAY, GLM, and MANOVA. ONEWAY and GLM are available using menus or syntax, and MANOVA is available only in syntax. GLM and MANOVA are more general than ONEWAY, which can only analyze single Between-S factor designs. If requested, the procedures provide all the statistics necessary to calculate the ANOVA.

```
ONEWAY psypath BY major /STATISTICS = DESCR.
```

	N	Mean	Std. Deviation
1.00	5	4.0000	2.54951
2.00	5	3.0000	2.00000
3.00	5	5.0000	2.12132
4.00	5	8.0000	1.00000
Total	20	5.0000	2.65568

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.000	3	23.333	5.833	.007
Within Groups	64.000	16	4.000		
Total	134.000	19			

The GLM results below contain the standard quantities computed earlier and just shown for ONEWAY, but they also contain other quantities. First, the SS in the Total line below equals 634.0, not

134.0, which appears on the Corrected Total line. Here GLM Total refers to the deviation of individual scores from 0; that is, GLM attempts to account for all variability relative to 0, rather than the grand mean. The value of 634.0 includes both variability due to the deviation of scores from the grand mean (i.e., SS_{Total} = 134.0) and deviations of the grand mean from 0 (i.e., the other 500.0 units). The variability due to the grand mean is shown on the Intercept line and is calculated by: $SS_{GrandMean} = N\times(\overline{y}_G - 0)^2 = 20\times(5.0 - 0)^2 = 500.0$.

GLM psypath BY major /PRINT = DESCR /PLOT = PROFILE(major).

major	Mean	Std. Deviation	N
1.00	4.0000	2.54951	5
2.00	3.0000	2.00000	5
3.00	5.0000	2.12132	5
4.00	8.0000	1.00000	5
Total	5.0000	2.65568	20

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Corrected Model	70.000(a)	3	23.333	5.833	.007	
Intercept	500.000	1	500.000	125.000	.000	***= N($\overline{y}_G$ – 0)^2***
major	70.000	3	23.333	5.833	.007	
Error	64.000	16	4.000			
Total	634.000	20				***= Σ(y – 0)^2 = 500.0 + 134.0***
Corrected Total	134.000	19				***= ANOVA SS_{Total}***

a R Squared = .522 (Adjusted R Squared = .433) ***= η^2 (eta^2) = SS_{Major}/SS_{Total}***

The other extra line is the Corrected Model line, which here is the same as the line for Major, the independent variable. The Corrected Model line in GLM (and the Model line in MANOVA) represent the overall effect of all factors in the design; that is, individual effects are added together. In the single-factor design, Model and factor lines are redundant.

The MANOVA command is similar to GLM except that the lowest and highest levels of the factor must follow the factor name; these values are usually 1 and k.

MANOVA psypath BY major(1 4) /PRINT = CELL.

FACTOR	CODE	Mean	Std. Dev.	N
major	1	4.000	2.550	5
major	2	3.000	2.000	5
major	3	5.000	2.121	5
major	4	8.000	1.000	5
For entire sample		5.000	2.656	20

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	64.00	16	4.00		
major	70.00	3	23.33	5.83	.007
(Model)	70.00	3	23.33	5.83	.007
(Total)	134.00	19	7.05		
R-Squared =	.522	Adjusted R-Squared =	.433		

The /PLOT = PROFILE(major) command in the preceding GLM produced the basic graph in Figure 1-1, which was then edited in SPSS's Chart Editor. Later follow-up analyses will permit more precise conclusions, but the graph and means in previous output indicate that the highest average psychopathy score was obtained for the Business students, as predicted by Hare's hypothesis and prior findings. Specifically, the mean for the five business students is higher than the means for each of the other three groups. Although H_0 that the population means are equal is rejected, the overall (i.e., omnibus) ANOVA does not permit conclusions about differences between specific pairs of means or between means grouped in various ways. These issues are addressed by follow-up analyses in the next two chapters.

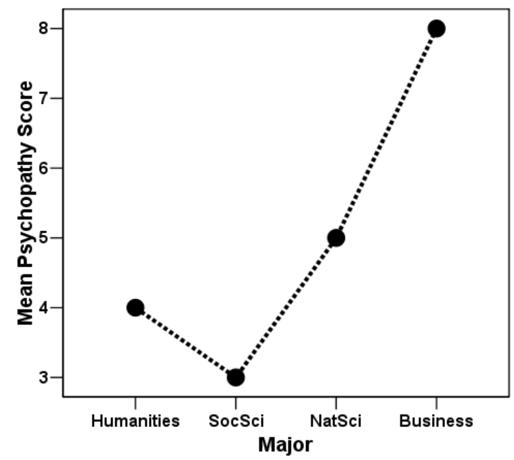


Figure 1-1. Plot of group means.

Using SPSS to Compute SSs for ANOVA

SPSS can calculate intermediate quantities required for the Between-S ANOVA. GLM is particularly helpful in this regard as it can vary what factor(s) are included in the analysis and save predicted and residual scores. The following variables created by GLM correspond to different values reported in the GLM output (e.g., *cortot* = GLM Corrected Total). Actual ANOVA results are not shown, as the purpose of the ANOVAs is solely to create these new variables.

The first GLM, which specifies no factor in the analysis, generates the Grand Mean as the predicted value for every subject (*grandmean*) and the deviation of individual scores from the grand mean (*cortot*). Deviations from the Grand Mean could also be created using COMPUTE.

The second GLM includes the major factor and generates Group Means as predicted scores (*groupmean*) and deviations of the scores from the Group Means as residuals (*error*).

The subsequent COMPUTE statements generate the deviation of each score from 0 (*glmtot*), the deviation of the Grand Mean from 0 (*glmint* for intercept), and the deviation of the Group Mean for each Major from the Grand Mean (*glmmaj*). These values are listed, squared in subsequent COMPUTE statements, and finally summed using the DESCRIPTIVE command.

```
GLM psypath /SAVE = PRED(grandmean) RESI(cortot).
...
GLM psypath BY major /SAVE = PRED(groupmean) RESI(error).
...
VARIABLE LABEL grandmean '' cortot '' groupmean '' error ''.
COMPUTE glmtot = psypath - 0.
COMPUTE glmint = grandmean - 0.
COMPUTE glmmaj = groupmean - grandmean.

LIST.
```

	y	$\bar{y}_G$	$y-\bar{y}_G$	$\bar{y}_j$	$y-\bar{y}_j$	$y-0$	$\bar{y}_G-0$	$\bar{y}_j-\bar{y}_G$
major	psypath	grandmean	cortot	groupmean	error	glmtot	glmint	glmmaj
1.0000	2.0000	5.0000	-3.0000	4.0000	-2.0000	2.0000	5.0000	-1.0000
1.0000	6.0000	5.0000	1.0000	4.0000	2.0000	6.0000	5.0000	-1.0000
1.0000	4.0000	5.0000	-1.0000	4.0000	.0000	4.0000	5.0000	-1.0000
1.0000	1.0000	5.0000	-4.0000	4.0000	-3.0000	1.0000	5.0000	-1.0000
1.0000	7.0000	5.0000	2.0000	4.0000	3.0000	7.0000	5.0000	-1.0000
2.0000	1.0000	5.0000	-4.0000	3.0000	-2.0000	1.0000	5.0000	-2.0000
2.0000	5.0000	5.0000	.0000	3.0000	2.0000	5.0000	5.0000	-2.0000
2.0000	3.0000	5.0000	-2.0000	3.0000	.0000	3.0000	5.0000	-2.0000
2.0000	1.0000	5.0000	-4.0000	3.0000	-2.0000	1.0000	5.0000	-2.0000
2.0000	5.0000	5.0000	.0000	3.0000	2.0000	5.0000	5.0000	-2.0000
3.0000	5.0000	5.0000	.0000	5.0000	.0000	5.0000	5.0000	.0000
3.0000	5.0000	5.0000	.0000	5.0000	.0000	5.0000	5.0000	.0000
3.0000	8.0000	5.0000	3.0000	5.0000	3.0000	8.0000	5.0000	.0000
3.0000	2.0000	5.0000	-3.0000	5.0000	-3.0000	2.0000	5.0000	.0000
3.0000	5.0000	5.0000	.0000	5.0000	.0000	5.0000	5.0000	.0000
4.0000	9.0000	5.0000	4.0000	8.0000	1.0000	9.0000	5.0000	3.0000
4.0000	8.0000	5.0000	3.0000	8.0000	.0000	8.0000	5.0000	3.0000
4.0000	7.0000	5.0000	2.0000	8.0000	-1.0000	7.0000	5.0000	3.0000
4.0000	7.0000	5.0000	2.0000	8.0000	-1.0000	7.0000	5.0000	3.0000
4.0000	9.0000	5.0000	4.0000	8.0000	1.0000	9.0000	5.0000	3.0000

```
*square the quantities.
COMPUTE glmtot2 = glmtot**2.
COMPUTE glmint2 = glmint**2.
COMPUTE cortot2 = cortot**2.
COMPUTE glmmaj2 = glmmaj**2.
COMPUTE error2  = error**2.

DESCR glmtot2 glmint2 cortot2 glmmaj2 error2 /STAT = SUM.
```

	N	Sum	
glmtot2	20	634.0000	***deviation of scores from 0***
glmint2	20	500.0000	***deviation of Grand Mean from 0***
cortot2	20	134.0000	***deviation of scores from Grand Mean***
glmmaj2	20	70.0000	***deviation of Group Means from Grand Mean***
error2	20	64.0000	***deviation of scores from Group Means***

Alternatively, we could get descriptive statistics for the computed variables before they are squared, and use $(n-1)s^2$ to obtain SSs. Given the following statistics, for example, $SS_{Major} = (20-1)1.91943^2 = 70.00$.

```
DESCR glmtot glmmaj error.
```

	N	Mean	Std. Deviation
glmtot	20	5	2.65568
glmmaj	20	.0000	1.91943
error	20	.0000	1.83533

$$y_{ij}-\bar{y}_G=(y_{ij}-\bar{y}_j)+(\bar{y}_j-\bar{y}_G)$$

Box 1-4. Partitioning SS_{Total}

The final *glmmaj* column demonstrates why the deviations of group means from the grand mean were multiplied by the number of observations in each group. In essence, the deviation of each score from the grand mean consists of two parts: the deviation of the score from the group mean and the deviation of the group mean from the grand mean, as shown algebraically in Box 1-4. Squaring and summing the three parts of this equation gives the formula for the single factor Between-S design shown previously and in Appendix 1-1. The last formula on line four is eta^2 and estimates the strength of the relationship. It is equivalent to R^2 in the preceding analyses.

APPENDIX 1-1: FORMULA FOR SINGLE FACTOR BETWEEN-S ANOVA

$$SS_{Total} = \sum_{i=1}^{k}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_G)^2 = (n-1)s_G^2 = SS_{Treatment} + SS_{Error}$$

$$SS_{Error} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(y_{ij}-\bar{y}_j)^2 = \sum_{j=1}^{k} SS_j = \sum_{j=1}^{k}(n_j-1)s_j^2 = SS_{Total} - SS_{T}$$

$$SS_{Treatment} = \sum_{i=1}^{k}\sum_{i=1}^{n_j}(\bar{y}_j-\bar{y}_G)^2 = \sum_{i=1}^{k} n_j(\bar{y}_j-\bar{y}_G)^2 = SS_{Total} - SS_{Er}$$

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$
H_a: One or more = False

$$MS_{Trt} = \frac{SS_{Trt}}{k-1} \qquad MS_{Err} = \frac{SS_{Err}}{N-k} \qquad F = \frac{MS_{Trt}}{MS_{Err}} \qquad \eta^2 = \frac{SS_{Trt}}{SS_{Total}}$$

CHAPTER 2 - POST HOC COMPARISONS FOR SINGLE FACTOR B-S DESIGNS

Omnibus ANOVAs that reject the null hypothesis and accept the alternative are ambiguous given more than two levels for the independent variable or factor because the omnibus *F* does not indicate which groups differ significantly. Moreover, even when the null hypothesis is not rejected, specific differences between groups could be significant by more focussed tests. Because of these limitations, the omnibus *F* is generally followed by multiple comparison tests to examine specific differences between groups. Follow-up analyses are of two sorts, post hoc (a posteriori) comparisons and planned (a priori) comparisons. The distinction is somewhat analogous to nondirectional and directional tests of significance.

Specifically, post hoc comparisons are conducted when researchers do not predict how groups will differ. In this case, multiple comparisons are performed without any prior expectations. The post hoc tests presented here perform all possible pairwise comparisons, although other post hoc tests can involve more complex comparisons (e.g., average of groups 1 and 2 versus average of groups 3 and 4). The number of pairwise comparisons (denoted c) is equal to k(k-1)/2. For example, c = 4(4-1)/2 = 6 pairwise comparisons for four groups: 12, 13, 14, 23, 24, 34. For k = 5, c = 5(5-1)/2 = 10 comparisons: 12, 13, 14, 15, 23, 24, 25, 34, 35, 45. Post hoc comparisons require that the omnibus *F* test be significant, which indicates that there is significant variability in the means. Most post hoc tests also require an adjustment for the fact that multiple comparisons are being made.

Planned or a priori comparisons are appropriate when researchers predict what pattern is expected *prior to seeing the results* (e.g., performance better for certain groups, scores will improve or deteriorate across levels of the factor). Because planned comparisons are predicted and usually involve fewer tests than post hoc procedures, the omnibus *F* need not be significant before carrying out planned comparisons and (arguably) little or no adjustment is required for the multiple tests being conducted. A case could even be made that the omnibus F does not need to be done given a priori comparisons.

This chapter covers four post hoc procedures: Least-Significant Difference (LSD), Student-Newman-Keuls (SNK), Tukey, and Bonferroni. These tests are ordered from a liberal test (LSD) that makes little or no adjustment for the multiple comparisons to a conservative test (Bonferroni) that involves a substantial adjustment. SNK and Tukey fall between LSD and Bonferroni tests. Rejecting the null for each comparison is easier for liberal tests, and more difficult for conservative tests.

Adjustments are generally recommended for multiple comparisons because the greater the number of statistical comparisons, the greater the likelihood that one or more comparison is significant by chance, resulting in a Type I error (i.e., rejection of a true null hypothesis). The probability of one or more Type I errors as a function of number of tests is 1 minus the probability of 0 Type I errors = $1 - (1 - \alpha)^c$, where c =

the number of comparisons and α is the probability of a Type I error for a single test. With four groups and α = .05, p(one or more Type I errors) = 1 - (1 - .05)^6 = 1 - .735 = .265, much higher than the probability of a Type I error for a single test. Adjusting for the number of comparisons (i.e., making it more difficult to reject H_0) reduces this overall error rate (sometimes called the Experimentwise error rate) and maintains a more acceptable probability of a Type I error. For a single test with two groups, c = 1 and 1 - (1 - α)^c = α.

LSD and Bonferroni Tests

One exception to adjusting for multiple comparisons is the LSD test, which is essentially an unadjusted set of t-tests. The only protection against Type I errors is the fact that the omnibus *F* test must be significant, giving a statistical reason to believe that the H_0 is false. The formula for calculation of the t-tests is shown in Box 2-1 and differs just slightly from the independent t-test between two groups.

$$t = \frac{\bar{y}_j - \bar{y}_{j'}}{\sqrt{MSE\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)}}$$

Box 2-1. T-test formula

The primary difference is that t_{LSD} uses MSE from the omnibus ANOVA rather than s_p^2 based on variability in just the two groups being compared. Using data for all groups provides a better estimate of the population variance. This also means that df_{LSD} = N - k, rather than $n_1 + n_2 - 2$ for the standard t-test, and that the denominator for all comparisons (i.e., the standard error of the difference between means) will be the same when n_j is the same for all groups. The use of j and j' as subscripts for the group means and n_j represents the fact that multiple ts will be calculated involving different groups (e.g., j = 1 and j' = 2, j = 1 and j' = 3, and so on). Here is the omnibus ANOVA for the psychopathy study.

```
ONEWAY psypath BY major /STATISTICS = DESCR.
```

	N	Mean	Std. Deviation	Std. Error
1.00	5	4.0000	2.54951	1.14018
2.00	5	3.0000	2.00000	.89443
3.00	5	5.0000	2.12132	.94868
4.00	5	8.0000	1.00000	.44721
Total	20	5.0000	2.65568	.59383

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.000	3	23.333	5.833	.007
Within Groups	64.000	16	4.000		
Total	134.000	19			

Box 2-2 shows calculations to compare means for groups 1 and 2. Other *t*s would be calculated in the same way with SE = 1.265. Given multiple t-tests, it helps to lay out post hoc results in a table as follows. Order means from low to high and arrange as rows and columns. Each cell above the diagonal (indicated by --- below) involves a comparison between two groups; here there are six cells, equal

$$t_{12} = \frac{4.0 - 3.0}{\sqrt{4.0(\frac{1}{5} + \frac{1}{5})}} = \frac{1.0}{1.265} = .79$$

Box 2-2. Calculation of t

to the number of pairwise comparisons. Cells below the diagonal are redundant and are ignored. Compute the *t*s and enter them as positive values in the appropriate cells. Column means minus row means will always be positive.

Results of *t* Calculations

Group		2	1	3	4	
	$\bar{y}_j$	3.0	4.0	5.0	8.0	
2	3.0	---	.79	1.58	3.95^{LB}	LSD $t_{Critical}$ = 2.120
1	4.0		---	.79	3.16^{LB}	Bonf $t_{Critical}$ = 3.008
3	5.0			---	2.37^{L}	
4	8.0				---	

Arranging post hoc tests in this manner ensures that all pairwise comparisons are calculated exactly once and also provides a calculation check in that the largest *t* will be in the upper right cell and *t*s become smaller moving left on the rows or down the columns. Ordered means also allows a concise summary of the results, as shown below. Another advantage discussed later is that the layout facilitates comparing observed statistics with critical values for the SNK test.

With df = 16, α = .05, and a nondirectional test because researchers do not have expectations for post hoc tests, $t_{Critical}$ = 2.120. This means that group 4 differs significantly from each of the other three groups, which do not differ significantly from one another. The superscript L in the above table indicates which comparisons are significant by the LSD procedure.

Summarizing the results of post hoc tests can be complicated, especially when numerous groups are involved. One way to communicate the results is to group means that do *not* differ significantly from one another as a set. Here, for example, means for groups 2, 1, and 3 are not significantly different from one another; that is, comparisons 21, 23, and 13 are all not significant. They constitute one set of three means. Means for these three groups are all significantly different from the mean for group 4, which constitutes a separate set by itself. To summarize, list the means according to size on a line and underline those that are not significantly different from one another, as follows. Means must be ordered from small to large for this to work. SPSS does this for some post hoc tests, but in a slightly different way.

LSD Results

Group	2	1	3	4
$\bar{y}_j$	3.0	4.0	5.0	8.0

Sometimes, a mean belongs to more than one set, as seen shortly for the Bonferroni procedure. For example, if groups 3 and 4 were not significantly different in the present study, then group 3 would belong to two sets: 213 and 34.

Although it may occasionally be legitimate to use the post-hoc LSD procedure, it makes no adjustment for multiple comparisons and is prone to inflated Type I errors. A very conservative method is the Bonferroni test. For Bonferroni, *t*s are tested for significance using α/c, where c is the number of comparisons. With 4 groups (i.e., 6 comparisons), the probability for each *t* must be less than .05/6 = .00833 before being called significant, using α=.05. Requiring a smaller p value for significance makes it more difficult to reject H_0, which reduces the probability of a Type I error. Note that $1\text{-}(1\text{-}.00833)^6 \approx .05$.

One challenge with Bonferroni is that tables of significance for *t* can only report a few discrete values for the area or probability (i.e., .10, .05, .025, and so on), rather than more specific values like .00833. On-line calculators and SPSS can compute critical values of *t* for any specified area; see Appendix 2-1, which includes an SPSS procedure to calculate observed p values for the Bonferroni test. The commands there show that the critical value of *t* for Bonferroni with 6 means is 3.008, considerably larger than the critical value for LSD. Bonferroni is more conservative in the sense that it is more difficult to reject the null hypothesis of no difference between the means. Making it more difficult to reject H_0 reduces the probability of rejecting a true H_0, that is, making a Type I error.

Some differences may be significant by the LSD test but not by Bonferroni. As shown in the table of observed *t*s above, only two differences represented by the subscript B are significant for Bonferroni versus three for LSD. An equivalent approach to the Bonferroni is in terms of p values rather than critical *t*s. This approach is based on the following equality and is used by SPSS:

$$\text{if } p_{\text{Observed}} < .05/c, \text{ then } c \times p_{\text{Observed}} < .05$$

That is, given a p value for the LSD test, multiply it times c, the number of comparisons, and then compare the resulting value to .05 (or whatever alpha is chosen). SPSS provides p values for post hoc tests, but SPSS can also compute these p values directly, as shown in the Appendix 2-1. The p values also appear in the SPSS output for the LSD procedure.

Obtaining post hoc results is straightforward in SPSS using either *ONEWAY* or *GLM*. The relevant subcommands are shown below. Because there is a single factor in ONEWAY, it is only necessary to specify the tests. GLM can have multiple factors, so users must specify what factor the post hoc tests are for. Note the correspondences between the following results and the preceding discussion: SE = 1.265, significance levels, p values, and Bonferroni $p = 6 \times p_{LSD}$. For example, Bonferroni $p_{14} = .036 = 6$ times LSD $p_{14} = .006$. When the Bonferroni adjustment produces a p > 1.000, SPSS prints 1.000 because a probability can never be greater than 1. Although SPSS does not compute final *t* values, it provides numerators (Mean Difference) and denominators (Std. Error) to compute *t*, if needed. Using menus to obtain these tests will show other post hoc tests that are available.

```
ONEWAY psypath BY major /RANGES = LSD BONFERRONI.
...
Post Hoc Tests
            (I)   (J)   Mean Difference Std.    Sig.
            major major (I-J)           Error
 LSD        1.00  2.00  1.00000         1.26491 .441
                  3.00  -1.00000        1.26491 .441
                  4.00  -4.00000(*)     1.26491 .006
            2.00  3.00  -2.00000        1.26491 .133
                  4.00  -5.00000(*)     1.26491 .001
            3.00  4.00  -3.00000(*)     1.26491 .031

 Bonferroni 1.00  2.00  1.00000         1.26491 1.000
                  3.00  -1.00000        1.26491 1.000
                  4.00  -4.00000(*)     1.26491 .036  t=4.0/1.26491=3.16>3.008
            2.00  3.00  -2.00000        1.26491 .800
                  4.00  -5.00000(*)     1.26491 .007
            3.00  4.00  -3.00000        1.26491 .184
* The mean difference is significant at the .050 level.
```

Note below how GLM specifies the factor then lists the desired post-hoc tests. The GLM commands produce the same output as ONEWAY, but can also be used in factorial studies with more than one factor.

```
GLM psypath BY major /POSTHOC = major(LSD BONF).
...
Post Hoc Tests
            (I)   (J)   Mean Difference Std.    Sig.
            major major (I-J)           Error
 LSD        1.00  2.00  1.0000          1.26491 .441
                  3.00  -1.0000         1.26491 .441
                  4.00  -4.0000(*)      1.26491 .006
...(same as ONEWAY results)
```

Summarizing the Bonferroni results must consider that four comparisons are not significantly different, 21 23 13 34, and two comparisons are significant, 24 14, a total of six comparisons. Specifically, group 3 must be included in two sets of means that are not significantly different, one set including 3 with 1 and 2, and the other set including 3 with 4. Here is the summary.

Bonferroni Results

Group	2	1	3	4
$\bar{y}_j$	3.0	4.0	5.0	8.0
	-------	-------	-------	
			-------	-------

The q (or Range) Statistic

$$q = \frac{\bar{y}_{Largest} - \bar{y}_{Smallest}}{\sqrt{MSE\left(\frac{1}{n_j}\right)}}$$

Box 2-3. Q-statistic

The Bonferroni test may be too conservative and reduce the probability of a Type I error so much that it inflates the probability of a Type II error excessively. Two procedures that fall between the LSD and Bonferroni tests are SNK (Student-Neuman-Keuls) and Tukey. Traditionally, the SNK and Tukey tests use the *q* or Range statistic shown in Box 2-3. Note that the denominators differ for *q* and *t*; specifically, the denominator for *q* is smaller, resulting in a larger value for *q* than for *t*. Therefore, the *q* statistic cannot be compared to critical values of *t*; rather it has its own table of critical values (see supplementary tables).

The sampling distribution (i.e., critical values for *q*) was obtained by randomly selecting k samples from a single population, where k is the number of samples selected, and calculating critical values for *q* with the formula in Box 2-3. The reasoning is that a larger k can elicit a larger $\overline{y}_{Largest} - \overline{y}_{Smallest}$ (the numerator for *q*) by chance. A difference is much more likely to be large with k = 7 than with k = 3, for example.

Figure 2-1 shows a simulation similar to how the Studentized Range statistic was initially calculated. From 2 to 8 samples of five observations were selected 10,000 times from the identical population and the difference between the largest and smallest sample means was calculated. As shown, the greater the number of samples the greater the difference between the largest and smallest means, despite both coming from the same population.

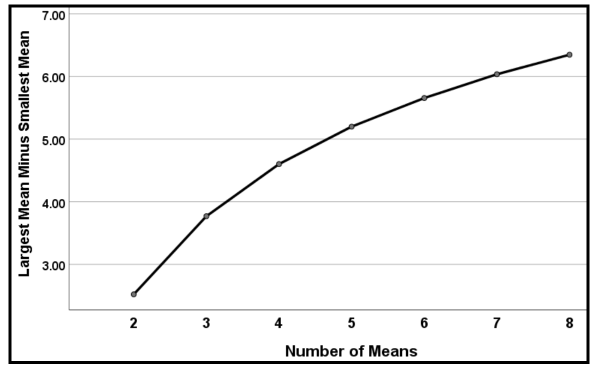


Figure 2-1. Simulation of Range Statistic

To accommodate this issue, the table for q includes a value called the Stretch, which represents the number of groups spanned by any particular comparison when group means are ordered from lowest to highest, including the two groups being compared. Comparing the largest and smallest means for only two groups, the span would be 2; comparing the largest and smallest means for three groups, the span would be 3; and so on.

Critical values of *q* become larger as Stretch increases so that the critical value to reject the null hypothesis increases as the chance probability of a large difference increases. With df = 16 and a = .05, $q_{Critical}$ = 3.00, 3.65, 4.05, and 4.33 for stretches 2 to 5, respectively. To carry out the SNK and Tukey post hoc tests, which differ in terms of how stretch is defined, *q* is calculated using the formula in Box 2-4 (the left formula is used when n_js are equal and the middle formula when n_js differ across samples).

$$q = \frac{\overline{y}_j - \overline{y}_{j'}}{\sqrt{MSE\left(\frac{1}{n_j}\right)}} = \frac{\overline{y}_j - \overline{y}_{j'}}{\sqrt{\frac{MSE}{2}\left(\frac{1}{n_j} + \frac{1}{n_{j'}}\right)}} = t\sqrt{2}$$

Box 2-4. Calculation of q statistic

The denominator of the middle version makes clear that the formulas for *q* and *t* differ because MSE is divided by 2 for *q*, which means that *q* will be √2 times *t*. Equivalently, *t* will equal q/√2. Given this relationship, when Stretch = 2, the critical value of q (3.00) is √2 times greater than the critical value for *t* (2.12); that is, $t_{Critical} \times \sqrt{2} = 2.12 \times \sqrt{2} = 2.998 \approx 3.00 = q_{Critical}$. Given the relationship between observed and critical values of *t* and *q* when stretch = 2, the *q* test with stretch = 2 is equivalent to a t-test. As stretch increases, the test becomes more conservative.

With equal n_js in the psychopathy study, the denominator for all the *q*s will be SQRT{4.0(1/5)} = .894 (which equals 1.265/√2, where 1.265 is the denominator calculated earlier for *t*). For group 2 versus group 1, $q = (4.0-3.0)/.894 = 1.12 = .79 \times \sqrt{2}$. The remaining *q*s appear in the following table.

Results of q Calculations

Group		*2*	*1*	*3*	*4*
	M_j	*3.0*	*4.0*	*5.0*	*8.0*
2	*3.0*	.	1.12	2.24	5.59^{LBT}
1	*4.0*		.	1.12	4.47^{LBT}
3	*5.0*			.	3.36^{L}
4	*8.0*				.

For df = 16, a = .05, and stretch = 2, $q_{Critical} = 3.00$, leading to the same result as the LSD test using *t*; that is, group 4 differs from groups 1, 2, and 3, which do not differ from one another. This is expected since the two tests are equivalent.

For the SNK and Tukey tests, the stretch used is generally greater than 2, leading to larger critical value for *q* and making the tests more conservative than LSD. The Tukey test uses Stretch = k, the number of groups. Hence, $q_{TukeyCritical} = 4.05$ when k = 4, which means that group 4 differs significantly from 2 and 1, but other differences are not significant, including that between groups 4 and 3. This is identical to the Bonferroni conclusion and is reflected in the superscript Ts in the above table.

Even the Tukey procedure may be too conservative because it uses k as the stretch for all comparisons. But when k means are ordered from smallest to largest, only the comparison between the most extreme means corresponds to Stretch = k. In the psychopathy study, for example, only the comparison between group 2 (lowest $\overline{y}_j$) and group 4 (highest $\overline{y}_j$) actually includes or spans all four groups, as shown below. For other comparisons, Stretch = k is too large.

2	1	3	4	Stretch	Comparison
-----------------------------				4	2 versus 4

Two comparisons span only three groups:

2	1	3	4		
--------------------				3	2 versus 3
	--------------------			3	1 versus 4

Three comparisons span only two groups:

2	1	3	4		
-----------				2	2 versus 1
	-----------			2	1 versus 3
		-----------		2	3 versus 4

The stretch for each cell is indicated below by a superscript. The three critical values are shown to the right of the observed *q*s. Comparing the observed and appropriate critical values indicates that, for the SNK tests, group 4 differs significantly from groups 2, 1, and 3, which do not differ significantly from one another. This duplicates the results for the LSD procedure and would be summarized in the same way.

Group		*2*	*1*	*3*	*4*		
	$\overline{y}_j$	*3.0*	*4.0*	*5.0*	*8.0*	Str	$q_{SNKCritical}$
2	*3.0*	.	1.12^2	2.24^3	5.59^{4SNK}	4	4.05
1	*4.0*		.	1.12^2	4.47^{3SNK}	3	3.65
3	*5.0*			.	3.36^{2SNK}	2	3.00
4	*8.0*				.		

Laying out the cells as we did simplifies comparisons between observed qs and SNK critical values. The top right cell will have a stretch of k, 4 in this study as shown by the superscript for that cell. The cells immediately to the left and below will have a stretch of k-1, 3 in the present case and identified by the superscript 3. This continues moving down and left until reaching the cells just above the diagonal, which have a stretch of 2 as indicated by the superscript 2. Or one can think of starting at the diagonal with a stretch of 2 and moving diagonally until reaching the top right cell with a stretch of k.

Because SNK uses different critical values, paradoxical outcomes are possible in that a smaller difference might be significant and a larger difference not significant. Imagine, for example, that the comparison between groups 4 and 1 produced $q_{Observed} = 3.50$ rather than 4.47. This value would not be significant compared to $q_{Critical} = 3.65$ whereas the smaller difference between group 4 and group 3 ($q_{Observed} = 3.36$) would be significant because of its lower $q_{Critical} = 3.00$. To avoid such anomalies, comparisons start with the largest difference (i.e., largest stretch and *q*) in the top right corner, working left across the row, and stopping for that row as soon as a difference is not significant. All comparisons to the left are considered not significant. Then the next row down would be examined in the same manner starting at the far right, but ***only if*** the cell above is significant. Using this procedure, cells to the left or below a non-significant cell are not significant. This procedure works when comparisons are arranged with rows and columns ordered from lowest to highest $\overline{y}_j$. Statistical packages, such as SPSS, incorporate this precaution into the SNK test.

All four post hoc procedures are shown below for GLM; the output would be the same for ONEWAY. The Tukey results have been added to the pairwise comparison section of the output, showing that comparisons 14 and 24 are significant. The SNK results are not shown in the pairwise comparison section because of possible paradoxical results. Instead, the SNK results, along with the Tukey results, appear in a new section of output labelled Homogeneous Subsets. This section is similar to the underlining procedure used to assign groups to sets that do not differ significantly from one another. The four groups have been ordered from low to high according to $\overline{y}_j$ and means that do not differ significantly appear in separate columns to the right. The SNK result corresponds to the underlining for the LSD procedure and the Tukey result corresponds to the underlining for the Bonferroni procedure because SNK and LSD produced identical conclusions as did Tukey and Bonferroni. The p values below each column show the most significant difference within the group; for example, p = .416 below Subset 1 for the Tukey test corresponds to the 23 comparison.

```
GLM psypath BY major /POSTHOC = major(LSD SNK TUKEY BONF).
...
Post Hoc Tests
            (I)   (J)   Mean Difference Std.    Sig.
            major major (I-J)           Error
 LSD        1.00  2.00  1.0000          1.26491 .441
                  3.00  -1.0000         1.26491 .441
                  4.00  -4.0000(*)      1.26491 .006
            2.00  3.00  -2.0000         1.26491 .133      Redundant lines deleted
                  4.00  -5.0000(*)      1.26491 .001            e.g.,2 vs 1, 3 vs 2
            3.00  4.00  -3.0000(*)      1.26491 .031

 Tukey HSD  1.00  2.00  1.0000          1.26491 .858
                  3.00  -1.0000         1.26491 .858
                  4.00  -4.0000(*)      1.26491 .028
            2.00  3.00  -2.0000         1.26491 .416
                  4.00  -5.0000(*)      1.26491 .006
            3.00  4.00  -3.0000         1.26491 .123

 Bonferroni 1.00  2.00  1.0000          1.26491 1.000
                  3.00  -1.0000         1.26491 1.000
                  4.00  -4.0000(*)      1.26491 .036      = 6×.006
            2.00  3.00  -2.0000         1.26491 .800
                  4.00  -5.0000(*)      1.26491 .007
            3.00  4.00  -3.0000         1.26491 .184

Homogeneous Subsets
                 major N Subset
                         1      2
 Student-Newman- 2.00  5 3.0000
 Keuls(a,b,c)    1.00  5 4.0000             Equivalent to:    2 1 3 4
                 3.00  5 5.0000                               -----
                 4.00  5        8.0000
                 Sig.     .282   1.000

 Tukey           2.00  5 3.0000
 HSD(a,b,c)      1.00  5 4.0000             Equivalent to:    2 1 3 4
                 3.00  5 5.0000 5.0000                        -----
                 4.00  5        8.0000                            ---
                 Sig.     .416   .123
```

To fully appreciate that the four tests from most liberal to most conservative are ordered LSD, SNK, Tukey, and Bonferroni, it helps to compare p values for each comparison across tests. Most ps can be extracted from the preceding tables, or SPSS can be used to generate them, as shown below (see Appendix 2-1). PLSD to PBON contain the p values and agree with those from the preceding analyses.

```
DATA LIST FREE /comp str tobs.
BEGIN DATA
12  2 .79   23 3 1.58   24 4 3.95   13 2  .79   14 3 3.16   34 2 2.37
END DATA.
COMP df = 16.
COMP c = 6.
COMP qobs = tobs*sqrt(2).

COMP plsd = 1 - CDF.SRANGE(qobs, 2, df).
COMP psnk = 1 - CDF.SRANGE(qobs, str, df).
COMP ptuk = 1 - CDF.SRANGE(qobs, 4, df).
COMP pbon = c*(1 - CDF.SRANGE(qobs, 2, df)).

FORMAT comp str df (F2.0) tobs qobs plsd psnk ptuk pbon (F6.4).
SORT CASES BY comp.
LIST.
```

comp	str	tobs	df	qobs	plsd	psnk	ptuk	pbon	
12	2	.7900	16	1.1172	.4411	.4411	.8579	2.6465	***LSD = SNK, str = 2***
13	2	.7900	16	1.1172	.4411	.4411	.8579	2.6465	***LSD = SNK, str = 2***
14	3	3.1600	16	4.4689	**.0061**	**.0158**	**.0279**	**.0364**	$p_{LSD} < p_{SNK} < p_{Tukey} < p_{Bonferroni}$
23	3	1.5800	16	2.2345	**.1337**	**.2825**	**.4168**	**.8020**	$p_{LSD} < p_{SNK} < p_{Tukey} < p_{Bonferroni}$
24	4	3.9500	16	5.5861	.0011	.0057	.0057	.0069	***SNK = TUK, str = k***
34	2	2.3700	16	3.3517	.0307	.0307	.1236	.1842	***LSD = SNK, str = 2***

Note that p values become larger (with some exceptions noted above) from PLSD to PBON, indicating that the comparisons become less significant, hence less likely to reject the null hypothesis and make a Type I error. Comparisons 14 and 23 show the increased conservatism best because no tests are equivalent. In general, LSD is more liberal than SNK, which is more liberal than Tukey, which is more liberal than Bonferroni, the most conservative test. Some p values are equal when stretches for the tests and hence the critical values are equal.

Post hoc comparisons are one of the more complex areas of statistics, largely because there are so many procedures (illustrated by options in the GLM post hoc menu in Figure 2-2). Essentially, researchers must decide about the cost of making a Type I error and hence how conservative to be for a particular study. For example, if the consequence of rejecting H_0 is costly or has other important implications, researchers should be more conservative. If the consequences of rejecting H_0 are minor or not very serious, then a more liberal test would be appropriate.

The results of post hoc procedures can produce anomalous (i.e., “messy”) results as illustrated above

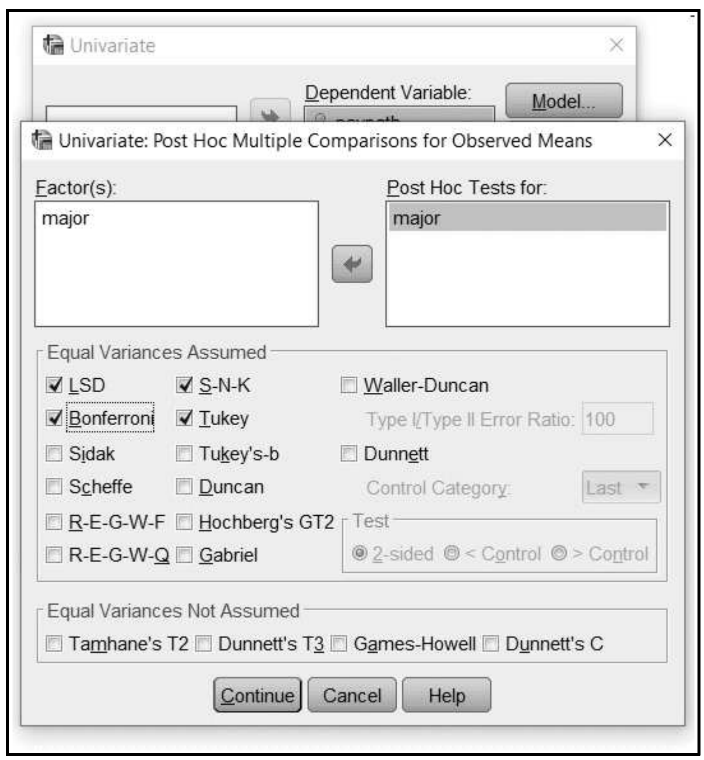


Figure 2-2. Post Hoc Menu for GLM.

for the Tukey and Bonferroni procedures. Specifically, 2=1=3 and 3=4, but 2≠4 and 1≠4. The results for LSD and SNK, on the other hand, lend themselves better to a more interpretable result. Specifically, 2=1=3 and 2≠4, 1≠4, and 3≠4, the Business students score higher than the three other Majors. Interpretable statistical results are more likely given well-founded predictions about the expected results and conducting planned or a priori tests, as discussed in chapter 3.

APPENDIX 2-1: USING SPSS TO CALCULATE CRITICAL TS AND P VALUES

The following commands compute the critical value of t using a Bonferroni value for alpha. The calculation of the critical value for LSD is also illustrated to show that it corresponds to the tabled value. The essential difference between the two computes is that .05 (equivalently .025 in each tail) is divided by c = 6 to obtain the critical t for Bonferroni, which is 3.0083. Changing the values for df and c in the data line would produce critical values for any study involving post hoc comparisons.

```
*critical values of t for LSD and Bonf.
*   df = 16, alpha = .05, nondir, c = #comp = 6.
DATA LIST FREE / df c.
BEGIN DATA
16 6
END DATA.

COMP tlsd  = IDF.T(1 - (.05/2), df).
COMP tbonf = IDF.T(1 - (.05/2)/c, df).

FORMAT df (F2.0) tlsd (F6.4) tbonf (F6.4).
LIST.
df   tlsd  tbonf

16 2.1199 3.0083
```

The following commands take a second approach and multiple the observed p value by the number of comparisons, 6 in this case. The resulting p can be compared to .05. The conclusions are the same irrespective of which approach is taken.

```
*p values for LSD and Bonf, df = 16, #comp = 6.
DATA LIST FREE /comp tobs.
BEGIN DATA
12  .79     23 1.58     24 3.95
            13  .79     14 3.16
                        34 2.37
END DATA.
COMP df = 16.
COMP c = 6.

COMP plsd = 2*(1 - CDF.T(tobs, df)).
COMP pbon = c*(2*(1 - CDF.T(tobs, df))).        could have been 6*plsd

FORMAT comp df (F2.0) plsd pbon (F6.4).
SORT CASES BY comp.
LIST comp tobs df plsd pbon.
comp   tobs df   plsd   pbon        LSD         Bonferroni
 12     .79 16  .4411 2.6465        ns          ns pbon>1
 13     .79 16  .4411 2.6465        ns          ns pbon>1
 14    3.16 16  .0061  .0364        sig         sig
 23    1.58 16  .1337  .8020        ns          ns
 24    3.95 16  .0011  .0069        sig         sig
 34    2.37 16  .0307  .1842        sig         ns
```

CHAPTER 3 - PLANNED COMPARISONS AND REGRESSION ANALYSES FOR SINGLE FACTOR BETWEEN-S DESIGNS

Post hoc tests are used when there are no expectations or predictions about the outcome of a study, but they do not always lead to tidy conclusions. In the psychopathy study, for example, the Tukey and Bonferroni tests conclude that groups 2, 1, and 3 did not differ from one another, that groups 3 and 4 did not differ, but groups 2 and 1 differed from group 4. The fact that group 3 is not significantly different from 2 and 1 or from 4 creates a problem for coming to a neat conclusion. In contrast, the LSD and SNK procedures in that particular study did lead to the tidy conclusion that the business students in group 4 obtained significantly higher psychopathy scores than all three groups from other faculties, which did not differ from one another.

A better approach for most studies (i.e., more likely to produce an interpretable pattern) is to make predictions about the expected results and perform tests or contrasts that correspond to those expectations, assuming the predictions are correct. An added benefit of such planned comparisons (called contrasts) is that the omnibus *F* need not be significant to carry out these follow-up analyses, a benefit because the omnibus *F* can be non-significant even when specific contrasts are significant. This can occur because $SS_{Between}$ is divided by k - 1 in the omnibus analysis even though most of the variability in $SS_{Between}$ could be due to just one comparison (i.e., df = 1).

There are also costs and risks associated with planned comparisons, however. One cost is that researchers usually restrict the number of tests to fewer comparisons than with post hoc procedures; one guideline for "fewer" is to limit the number of comparisons to the *df* for the factor (i.e., number of comparisons is c = k - 1). A second risk is that the predictions might be wrong, leading to analyses that do not in fact correspond well to the observed data. The availability of prior studies and well-founded theories determine how great that risk is. But science does often benefit from rejected hypotheses and theories!

Planned comparisons make use of ***contrasts*** (also called linear contrasts), which essentially are k coefficients (numbers), one for each of the k groups. These coefficients (denoted by $c_j = c_1, c_2, \ldots c_k$) test the significance of patterns expected in the data and must sum to 0. With k = 4, the following are some possible patterns (labels in the left column are explained shortly):

Group	c_1	c_2	c_3	c_4	Σc_j
C_{12v34}	-1	-1	+1	+1	0
C_{linear}	-3	-1	+1	+3	0
$C_{nonlinear}$	-1	+1	+1	-1	0
C_{123v4}	-1	-1	-1	+3	0
C_{12v3}	-1	-1	+2	0	0
C_{1v2}	-1	+1	0	0	0

Contrast analyses essentially tests whether these predicted patterns correlate strongly enough with the observed cell means to reject the null hypothesis of no relationship. Some contrasts can be interpreted as differences between $\overline{y}_j$s as in C12v34, C123v4, C12v3, and C1v2 above. Specifically, $\overline{y}_j$s for groups with negative numbers are being contrasted with $\overline{y}_j$s for groups with positive numbers. For example, C_{123v4} tests whether the average for groups 1, 2, and 3 differs significantly from the average for group 4. Other contrasts define more complex patterns or differences. For example, the coefficients for C_{linear} increase in a linear fashion from group 1 to group 4. A systematic increase or decrease in means would correlate positively or negatively with these coefficients. $C_{nonlinear}$ defines a curvilinear or nonlinear pattern. Means that showed a U shaped or inverted U shape pattern would correlate with the $C_{nonlinear}$ coefficients. Whether the correlation is positive or negative is incidental to the statistical test, although critical for interpretation of the results.

A common practice is to limit tests to k - 1 contrasts with the added restriction that contrasts are orthogonal to one another; orthogonal means independent or uncorrelated. Specifically, two sets of contrast coefficients, c_j and c_j', are orthogonal if $\Sigma c_j c_j' = 0$, that is, if the cross products of the coefficients sum to 0. The lack of correlation given this condition occurs because SCP = 0 if $\Sigma c_j c_j' = 0$ for contrasts (the computational formula for SCP = Σxy - (ΣxΣy)/n and for contrasts Σx = 0 and Σy = 0). The test of orthogonality is illustrated below for several pairs shown above.

						$\Sigma c_j c_j'$	
C_{12v34}	-1	-1	+1	+1			
C_{linear}	-3	-1	+1	+3			
$C_{12v34} \times C_{linear}$	+3 +	+1 +	+1 +	+3	=	8	NOT orthogonal
C_{linear}	-3	-1	+1	+3			
$C_{nonlinear}$	-1	+1	+1	-1			
$C_{linear} \times C_{nonlinear}$	+3 +	-1 +	+1 +	-3		0	Orthogonal
C_{123v4}	-1	-1	-1	+3			
C_{12v3}	-1	-1	+2	0		0	Orthogonal

C_{12v3}	-1	-1	+2	0		
C_{1v2}	-1	+1	0	0	0	Orthogonal
C_{123v4}	-1	-1	-1	+3		
C_{1v2}	-1	+1	0	0	0	Orthogonal

Summing the products of contrast coefficients determines whether they are orthogonal. The challenge, however, is generating k - 1 orthogonal contrasts, although some common patterns emerge with practice. The third pair, for example, is orthogonal because C_{12v3} compares groups 1 and 2 with 3, and these three groups were all coded -1 (i.e., the same) in the preceding contrast, C_{123v4}. Therefore, C_{123v4} did not capture any variability in the means for groups 1, 2, and 3.

Three of the above contrasts are mutually orthogonal and correspond to meaningful comparisons for the psychopathy study. C_{123v4} compares the mean for groups 1, 2, and 3 with the mean for business students (i.e., the group predicted to score higher on psychopathy). This is the primary prediction. C_{12v3} compares humanities and social science students with natural science students, and C_{1v2} compares humanities students with social science students. These latter two "hypotheses" complete the required k - 1 orthogonal contrasts, but may be of less theoretical interest than C_{123v4}. In many studies only some of the k-1 contrasts are of interest.

$$L = \sum_{j=1}^{k} c_j \bar{y}_j \qquad t_L = \frac{L - 0}{\sqrt{MSE \sum_{j=1}^{k} \frac{c_j^2}{n_j}}}$$

$$SS_L = \frac{n_j L^2}{\sum_{j=1}^{k} c_j^2} \qquad F_L = \frac{SS_L/1}{MSE} \qquad df = 1, N-k$$

Box 3-1. Formula for planned comparisons

Given k-1 contrasts, the next step is to calculate a contrast score (L) for each contrast. As shown in Box 3-1, L is the sum of the cross products of the contrast coefficients and corresponding means. L represents the "correlation" between the contrast coefficients and the means, and can be tested for significance using a *t* test or an equivalent *F* test. Box 3-1 shows relevant formula, where MSE is MS_{Error} from the omnibus ANOVA. Calculations for the psychopath study appear below. $\Sigma c_j^2 = -1^2 + -1^2 + -1^2 + 3^2 = 12$ in denominator for SS_L.

	1-Hum	*2-SS*	*3-NS*	*4-Bus*		
$\bar{y}_j$	4.0	3.0	5.0	8.0	L	SS_L
C123v4	-1	-1	-1	+3	+12.0	$60.0 = 5 \times 12.0^2 / 12$
C12v3	-1	-1	+2	0	+ 3.0	7.5
C1v2	-1	+1	0	0	- 1.0	2.5
						$\Sigma SS_L = 70.0 = SS_{Major}$

To illustrate the above calculations for the first contrast, C_{123v4}:

$$L = -1\times4.0 + -1\times3.0 + -1\times5.0 + 3\times8.0 = +12.0$$

When k - 1 orthogonal contrasts are used, as above, the sum of the SS_Ls equals $SS_{Between}$ from the omnibus ANOVA. $SS_{Between}$ is partitioned into k - 1 components, one associated with each contrast. Each contrast represents one *df* from the omnibus *F*. Of particular note, most of the 70.0 units of variability loaded on our first contrast, as predicted, whereas the omnibus ANOVA divided 70.0 into k - 1 = 3 equal units.

$$t = \frac{12.0 - 0}{\sqrt{4.0\left(\frac{-1^2}{5} + \frac{-1^2}{5} + \frac{-1^2}{5} + \frac{+3^2}{5}\right)}} = \frac{12.0}{3.098} = 3.873$$

$$F = \frac{\frac{60.0}{1}}{4.0} = 15.0 = 3.873^2$$

Box 3-2. Tests of significance for psychopath study

` Both L and SS_L can be used to test significance. The conclusions for *t* and *F* will be equivalent. Box 3-2 shows calculations for the first contrast, c_{123v4}. The *t*-test (*df* = N - k) and *F*-test (*df* = 1, N - k) are equivalent because $df_{Numerator} = 1$ for the F test.

Each contrast shown here corresponds to a difference between means and the *t* and *F* statistics just calculated using L agree with *t* and *F* tests studied earlier for differences between means. Box 3-3 illustrates for contrast C_{123v4}. The strength of the contrast approach is that it also works for patterns (e.g., linear or nonlinear) that do not represent simple differences between means.

$$t_{123v4} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{8.0 - \frac{4.0 + 3.0 + 5.0}{3}}{\sqrt{4.0\left(\frac{1}{15} + \frac{1}{5}\right)}} = \frac{8.0 - 4.0}{1.032796} = 3.873 = t_{L1}$$

$$SS_{123v4} = 15(4.0 - 5.0)^2 + 5(8.0 - 5.0)^2 = 15 \times -1.0^2 + 5 \times 3.0^2 = 60.0 = SS_{L1}$$

Box 3-3. Contrast tests as difference between means.

Planned Contrasts in SPSS

ONEWAY, GLM, and MANOVA all provide ways to conduct planned contrasts. ONEWAY produces results shown below. Two *t*s are reported, one that assumes equal variances and a second that does not make that assumption. The first *t* shows various correspondences with our preceding analysis: L = Value of Contrast, SE, *t*, and *df*. The contrast is significant, even by a nondirectional test, although planned contrasts generally involve directional predictions. ONEWAY does not partition $SS_{Between}$.

```
ONEWAY psypath BY major /CONTRAST = -1 -1 -1 3.
                Sum of Squares df Mean Square F     Sig.
 Between Groups 70.000          3  23.333      5.833 .007
 Within Groups  64.000         16  4.000
 Total          134.000        19

                        Contrast Value of   Std. Error t     df     Sig.
                                 Contrast                           (2-tailed)
  psypath Assumes equal  1        12.0000    3.09839    3.873 16     .001
          variances
         Does not assume 1        12.0000    2.19089    5.477 14.445 .000
         equal variances
```

The following analyses illustrate features of the GLM procedure; the factor name is specified with the /CONTRAST command, which is followed by numerical values for one or more contrasts in parentheses after the SPECIAL option. SPSS also allows keywords for built-in contrasts described later. The first section of output shows the significance along with values that can be used to calculate *t* (i.e., numerator, denominator) but not the actual *t*. The following ANOVA summary table includes SS_L, F_L, and significance. The *t* and *F* agree with earlier calculations and correspond to one another; ps are equal and $F = t^2$.

```
GLM psypath BY major /CONTRAST(major) = SPECIAL(-1 -1 -1 3).
...
Custom Hypothesis Tests
 major Special                                 Dependent
 Contrast                                      Variable
                                               psypath
 L1            Contrast Estimate               12.000      t = 12.0/3.098 = 3.873
               Hypothesized Value              0
               Std. Error                      3.098
               Sig.                            .001

 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 60.000         1  60.000      15.000 .001   F = t2
 Error    64.000         16 4.000
```

The next analysis shows SPECIAL followed by k - 1 = 3 sets of k coefficients that correspond to the three contrasts. GLM produces results for three tests of significance, again providing quantities for the *t* statistic, but only a single ANOVA table that aggregates SSs for the three contrasts. The aggregate ANOVA corresponds to the omnibus ANOVA and adds nothing new. Only the first contrast comparing business students to the three other groups is significant, a tidy outcome consistent with the theoretical prediction.

```
GLM psypath BY major /CONTRAST(major) = SPECIAL(-1 -1 -1 3  -1 -1 2 0  -1 1 0 0).
...
Custom Hypothesis Tests
 L1            Contrast Estimate               12.000
               Std. Error                      3.098
               Sig.                            .001

 L2            Contrast Estimate               3.000
               Std. Error                      2.191
               Sig.                            .190

 L3            Contrast Estimate               -1.000
               Std. Error                      1.265
               Sig.                            .441

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 70.000         3  23.333      5.833 .007
 Error    64.000         16 4.000
```

The following GLM specifies three separate CONTRAST subcommands, one for each contrast. This produces a separate ANOVA for each contrast, illustrating that $SS_{Treatment}$ has been partitioned into three independent components. Note that the *p* value for contrast 3 equals the corresponding LSD result in chapter

2 because this is a pairwise comparison (1 vs 2) and no adjustment has been made.

```
GLM psypath BY major
 /CONTRAST(major) = SPECIAL(-1 -1 -1  3)
 /CONTRAST(major) = SPECIAL(-1 -1  2  0)
 /CONTRAST(major) = SPECIAL(-1  1  0  0).
...
Custom Hypothesis Tests #1
 L1            Contrast Estimate           12.000
               Std. Error                  3.098
               Sig.                        .001
 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 60.000         1  60.000      15.000 .001
 Error    64.000         16 4.000

Custom Hypothesis Tests #2
 L1            Contrast Estimate           3.000
               Std. Error                  2.191
               Sig.                        .190
 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 7.500          1  7.500       1.875 .190
 Error    64.000         16 4.000

Custom Hypothesis Tests #3
 L1            Contrast Estimate           -1.000
               Std. Error                  1.265
               Sig.                        .441
 Source   Sum of Squares df Mean Square F    Sig.
 Contrast 2.500          1  2.500       .625 .441
 Error    64.000         16 4.000
```

A second way to perform contrasts with GLM is the LMATRIX option, which is useful for analyses of factorial designs. The results are identical to those obtained with CONTRAST.

```
GLM psypath BY major /LMATRIX major -1 -1 -1 3.
...
Custom Hypothesis Tests
 L1       Contrast Estimate           12.000
          Std. Error                  3.098
          Sig.                        .001

 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 60.000         1  60.000      15.000 .001
 Error    64.000         16 4.000
```

MANOVA has many options for how planned contrasts are requested and how they appear in the output. One feature of MANOVA is that k - 1 orthogonal contrasts must be specified. The first MANOVA shows how to specify contrasts. The format is similar to GLM but the k - 1 orthogonal sets of k coefficients within the SPECIAL brackets follow k 1s that are entered first (these represent the grand mean). Default *t*-test results follow the omnibus ANOVA and equal those observed previously. These same *t* tests are also reported in later MANOVAs, but are deleted as redundant.

```
MANOVA psypath BY major(1 4)
   /CONTRAST(major) = SPECIAL(1 1 1 1  -1 -1 -1 3   -1 -1 2 0  -1 1 0 0).
...
 Estimates for psypath --- Individual univariate .9500 confidence intervals
  Parameter         Coeff.          Std. Err.          t-Value          Sig. t
        2      12.0000000000          3.09839          3.87298          .00135
        3       3.0000000000          2.19089          1.36931          .18982
        4      -1.0000000000          1.26491          -.79057          .44076
```

The next MANOVA includes a /PRINT SIGNIFICANCE(SINGLEDF) subcommand that requests SPSS to partition every Between-S effect with *df* > 1 (i.e., for factors with k > 2) into single *df* (SINGLEDF) tests of significance, using the contrasts specified in the CONTRAST subcommand (or default contrasts if none are specified). The overall SS_{Major} is partitioned on the three lines after the omnibus F. The analysis shows that the omnibus *F* for a Between-S factor is the average of the single *df Fs*; that is, (15.00 + 1.87 + .63)/3 = 5.833 = F_{Major}. That large and small *F*s are averaged together to produce the omnibus *F* demonstrates that an omnibus *F* can fail to be significant and a specific planned contrast significant if it captures enough variability among the means with its df = 1.

```
MANOVA psypath BY major(1 4) /PRINT = SIGNIFICANCE(SINGLEDF)
   /CONTRAST(major) = SPECIAL(1 1 1 1  -1 -1 -1 3   -1 -1 2 0  -1 1 0 0).

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              64.00      16      4.00
 major                     70.00       3     23.33      5.83      .007
   1ST Parameter           60.00       1     60.00     15.00      .001
   2ND Parameter            7.50       1      7.50      1.87      .190
   3RD Parameter            2.50       1      2.50       .63      .441

 (Model)                   70.00       3     23.33      5.83      .007
 (Total)                  134.00      19      7.05
...
```

The final MANOVA illustrates how to request single *df F* tests using the /DESIGN option. The default for MANOVA is /DESIGN factorname (*major* for the psychopathy study), which was omitted in preceding analyses. Instead of an overall major effect with *df* = 3, the following /DESIGN statement asks for three separate components of the major effect; *major*(1) denotes the first contrast, *major*(2) the second contrast, and *major*(3) the third contrast. The numbers in parentheses represent the k - 1 = 3 contrasts, *not* the k = 4 levels for major.

```
MANOVA psypath BY major(1 4)
   /CONTRAST(major) = SPECIAL(1 1 1 1  -1 -1 -1 3   -1 -1 2 0  -1 1 0 0)
   /DESIGN major(1) major(2) major(3).

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           64.00      16      4.00
 MAJOR(1)                  60.00       1     60.00     15.00      .001
 MAJOR(2)                   7.50       1      7.50      1.87      .190
 MAJOR(3)                   2.50       1      2.50       .63      .441
```

Appendix 3-1 summarizes the various ways to request contrasts in SPSS.

Regression Analyses for the Between-S Single Factor Design

Simple regression with a single predictor can conduct ANOVA for differences between two groups, but it can be extended to k > 2 with multiple regression. The basic principle is that *p*, the number of predictors, must equal k - 1. That is, for two groups, *p* = 1; for three groups, *p* = 2; and so on. How the k - 1 predictors are created can vary. One approach is the *p* predictors *can* (but need not) correspond to k - 1 orthogonal contrasts. The first step is to create three predictors, one corresponding to each contrast; these predictors are created using RECODE statements and appear in a later listing of the data file. Then the dependent variable *psypath* is regressed on these three predictors. Entering *c123v4* second means that CHANGE statistics will reflect the contribution of that contrast.

```
RECODE major (1 2 3 = −1) (4 = 3)           INTO c123v4.    −1 −1 −1 +3
RECODE major (1 2   = −1) (3 = 2) (4 = 0)   INTO c12v3.     −1 −1 +2  0
RECODE major (1     = −1) (2 = 1) (3 4 = 0) INTO c1v2.      −1 +1  0  0

REGRESS /STAT = DEFAULT CHANGE /DESCR /DEP = psypath
   /ENTER c12v3 c1v2 /ENTER c123v4   /SAVE PRED(prdp.m) RESI(resp.m).
```

```
         Mean   Std. Deviation N
 psypath 5.0000 2.65568        20
 c123v4  .0000  1.77705        20           Predictors are contrasts, Ms = 0
 c12v3   .0000  1.25656        20
 c1v2    .0000  .72548         20

                     psypath c123v4 c12v3
             c123v4  .669
             c12v3   .237    .000               Predictors are orthogonal rs = 0
             c1v2    −.137   .000   .000
```

The preliminary descriptive statistics reveal some interesting aspects of this analysis. First, *M*s = 0 for the predictors because contrast coefficients sum to 0. Also, *r*s = 0 for all three correlations between predictors because the three contrasts are orthogonal. Also each contrast has a non-zero *r* with the dependent variable, although *r* = 0 is possible for contrasts that are independent of the means. Squaring the *r*s and multiplying by SS_{Total} matches earlier calculations for the SS_Ls. To illustrate, $.669^2 \times 134.00 = 59.97 \approx 60.00$ = SS for the C_{123v4} contrast.

The remainder of the analysis reproduces earlier ANOVAs. The overall ANOVA table for Model 2 with all three predictors corresponds to the omnibus *F* because the predicted values are the group means (see listing below) and the residual values equal y - $\overline{y}_j$, as for the MSE in ANOVA.

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics R Square Change	F Change	df1	df2	Sig. F Change
1	.273(a)	.075	−.034	2.70076	.075	.685	2	17	.517
2	.723(b)	.522	.433	2.00000	.448	**15.000**	**1**	**16**	**.001**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10.000	2	5.000	.685	.517(a)
	Residual	124.000	17	7.294		
	Total	134.000	19			

```
 2     Regression 70.000          3  23.333       5.833 .007(b)     omnibus F
       Residual   64.000          16 4.000
       Total      134.000         19

 Model            Unstandardized             Standardized    t      Sig.
                  Coefficients               Coefficients
                  B               Std. Error Beta
...
 2     (Constant) 5.000           .447                       11.180 .000
       c12v3      .500            .365       .237            1.369  .190
       c1v2       -.500           .632       -.137           -.791  .441
       c123v4     1.000           .258       .669            3.873  .001
```

With respect to individual predictors, the results for the regression coefficients correspond to earlier calculations and output from ONEWAY, GLM, and MANOVA, including: the *t* and *p* values for the three contrasts, and F_{Change} for c123v4. Also, $SS_{Change} = 70.0 - 10.0 = 60.0 = SS_L$ for c_{123v4}.

Here is the data file for the study. Only the first two subjects in each group are shown.

```
VARIABLE LABEL prdp.m '' resp.m ''.
LIST.
```

```
           y                                  ȳj          y - ȳj
 major psypath c123v4  c12v3   c1v2       prdp.m      resp.m
  1.00    2.00  -1.00  -1.00  -1.00      4.00000    -2.00000
  1.00    6.00  -1.00  -1.00  -1.00      4.00000     2.00000
...
  2.00    1.00  -1.00  -1.00   1.00      3.00000    -2.00000
  2.00    5.00  -1.00  -1.00   1.00      3.00000     2.00000
...
  3.00    5.00  -1.00   2.00    .00      5.00000      .00000
  3.00    5.00  -1.00   2.00    .00      5.00000      .00000
...
  4.00    9.00   3.00    .00    .00      8.00000     1.00000
  4.00    8.00   3.00    .00    .00      8.00000      .00000
...
```

One advantage of multiple regression is that one or more numerical predictors could be included to determine the effect of controlling for predictors on differences between the means. Numerical predictors can also reduce MSE.

Predictors need not be orthogonal, as for contrasts, because regression can determine the unique contribution of each predictor. It is therefore possible to create predictors that correspond to some (but not all six) of the pairwise comparisons tested by the post hoc procedures. The following analysis compares business students (group 4) to each of the other groups in turn. Note in the descriptive statistics that the predictors do not correspond to contrasts (Ms $\neq 0$) and are not orthogonal (rs = -.333 $\neq$ 0). The overall ANOVA remains the same because predicted scores are group means. Tests for individual predictors agree with pairwise comparisons from earlier; compare differences between means, SEs, ts, and ps from the following regression and earlier post-hoc analyses.

```
RECODE major (4 2 3 = 0) (1 = 1) INTO c4v1.
RECODE major (4 1 3 = 0) (2 = 1) INTO c4v2.
RECODE major (4 1 2 = 0) (3 = 1) INTO c4v3.
```

```
REGRESS /DESCR /DEP = psypath /ENTER c4v1 c4v2 c4v3
 /SAVE PRED(prdp.m2) RESI(resp.m2).
```

	Mean	Std. Deviation	N
psypath	5.0000	2.65568	20
c4v1	.2500	.44426	20
c4v2	.2500	.44426	20
c4v3	.2500	.44426	20

Predictors not contrasts, Ms ≠ 0

	psypath	c4v1	c4v2
c4v1	-.223		
c4v2	-.446	-.333	
c4v3	.000	-.333	-.333

Predictors not orthogonal, rs ≠ 0

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.723(a)	.522	.433	2.00000

Model		Sum of Squares	df	Mean Square	F	Sig.
1	**Regression**	**70.000**	**3**	**23.333**	**5.833**	**.007(a)**
	Residual	**64.000**	**16**	**4.000**		
	Total	**134.000**	**19**			

Omnibus ANOVA

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	8.000	.894		8.944	.000
	c4v1	**-4.000**	**1.265**	-.669	**-3.162**	**.006**
	c4v2	**-5.000**	**1.265**	-.836	**-3.953**	**.001**
	c4v3	**-3.000**	**1.265**	-.502	**-2.372**	**.031**

```
VARIABLE LABEL prdp.m2 '' resp.m2 ''.
LIST major psypath c4v1 c4v2 c4v3 prdp.m2 resp.m2.
```

major	psypath	c4v1	c4v2	c4v3	prdp.m2	resp.m2
1.00	2.00	1.00	.00	.00	4.00000	-2.00000
1.00	6.00	1.00	.00	.00	4.00000	2.00000
...						
2.00	1.00	.00	1.00	.00	3.00000	-2.00000
2.00	5.00	.00	1.00	.00	3.00000	2.00000
...						
3.00	5.00	.00	.00	1.00	5.00000	.00000
3.00	5.00	.00	.00	1.00	5.00000	.00000
...						
4.00	9.00	.00	.00	.00	8.00000	1.00000
4.00	8.00	.00	.00	.00	8.00000	.00000

SPSS Menu System and Alternative Types of Contrasts

Although some contrasts correspond to differences between means, that is not the case for all contrasts. Polynomial contrasts, for example, are orthogonal contrasts that partition $SS_{Treatment}$ into linear, quadratic, cubic, and so on components, with each component representing a linear or nonlinear pattern. For k = 3, the linear coefficients are -1 0 +1 and the quadratic coefficients are 1 -2 1. The former capture a consistent increase or decrease in the means, whereas the latter capture the curvilinear pattern possible for three groups, either U-shaped or inverted U-shaped. For k = 4, the linear coefficients are -3 -1 1 3, the quadratic are 1 -1 -1 1, and the cubic are -1 3 -3 1. Note the increasing number of changes in direction (bends) necessary to capture all variability in means as k increases. These and polynomial coefficients for larger values of k can be found in the tables.

Users do not always have to specify numerical coefficients for contrasts. Both MANOVA and GLM have built-in coefficients for certain common contrasts, including contrasts presented earlier (called Difference contrasts) and Polynomial contrasts. The MANOVA analysis on the next page requests polynomial contrasts for the psychopathy study. The linear and quadratic components are significant, indicating that psychopathy increases significantly from Humanities to Social Science to Natural Science to Business majors, but the means deviate significantly from a pure linear relationship. The linear contrast, the most significant, accounts for 49.0 units of variability, whereas our earlier first contrast (-1 -1 -1 3) accounted for 60.0 units. That is, the earlier contrast provided a better single df fit to the data.

Figure 3-1 shows the selection of contrasts from the GLM Menu. After specifying the overall design, the Contrasts option is selected and brings up the top option screen. For each factor in the design, it is possible to select one of the available contrast types and click on Change to insert it after the factor name. The default contrast for each factor is None. Difference contrasts (also called Reverse Helmert contrasts) correspond to the preceding contrasts for the psychopathy study (i.e., -1 -1 -1 3, -1 -1 2 0, -1 1 0 0). The final syntax includes only the label for the contrast type and all k - 1 contrasts are carried out.

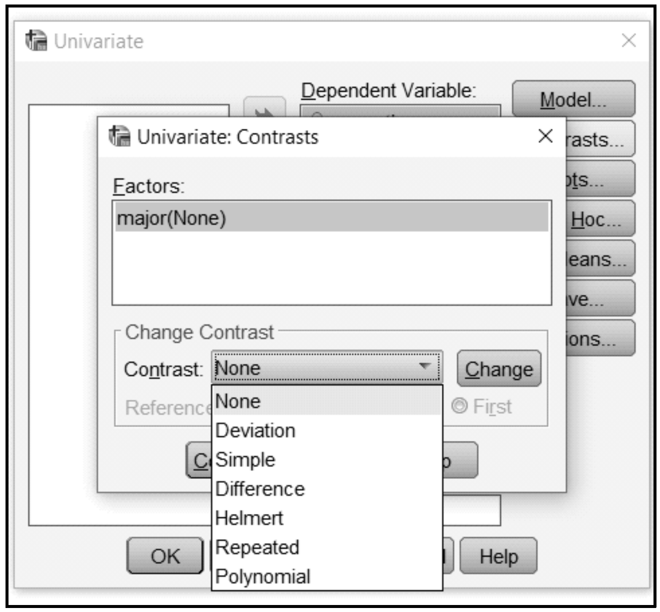


Figure 3-1. GLM Contrast Menu.

One warning about SPSS contrasts. GLM and MANOVA sometimes use different numerical values for contrast coefficients that reflect the same pattern as the integer values we use. In these cases, the final statistics (e.g., *t*, *F*, SS) will agree with calculations, but some intermediate values may not. Previously for example, manual calculations for the linear contrast shown below produced L = -3×4.0 + -1×3.0 + 1×5.0 +3×8.0 = 14.0, which gave $SS_{Linear} = 5 \times 14.0^2/20 = 49.0$. The linear coefficients used by SPSS were normalized: -.6708, -.2236, +.2236, +.6708, which gives $\Sigma c_j \times \overline{y} = 3.1305$, as shown below. Squared normalized coefficients sum to 1.0 (i.e., $\Sigma c_j^2 = 1.0$). To calculate normalized coefficients, each coefficient is divided by the square root of the sum of the integer coefficients squared. For example, $nc_{linear} = -3 \div \sqrt{20} = -.6708$, where $20 = -3^2 + -1^2 + 1^2 + 3^2$. Because $\Sigma c_j^2 = 1.0$, $SS_{Linear} = 5 \times 3.1305^2/1 = 49.0$.

```
MANOVA psypath BY major(1 4) /PRINT = SIGNIF(SINGLE)
   /CONTRAST (major) = POLYNOMIAL.
```

```
 Source of Variation            SS      DF        MS         F  Sig of F
 WITHIN CELLS                64.00      16      4.00
 major                       70.00       3     23.33      5.83      .007
   1ST Parameter             49.00       1     49.00     12.25      .003
   2ND Parameter             20.00       1     20.00      5.00      .040
   3RD Parameter              1.00       1      1.00       .25      .624

 (Model)                     70.00       3     23.33      5.83      .007
 (Total)                    134.00      19      7.05

 Estimates for psypath  --- Individual univariate .9500 confidence intervals
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper
        2   3.13049517     .89443    3.50000     .00296    1.23439    5.02660
        3   2.00000000     .89443    2.23607     .03994     .10390    3.89610
        4   -.44721360     .89443    -.50000     .62388   -2.34331    1.44889
```

Conclusions

Appendix 3-2 briefly presents some research outcomes that can be used to practice contrasts. This finishes single factor Between-S ANOVA. The next unit covers research designs with two Between-S factors, including the default ANOVA, post-hoc comparisons, and planned comparisons. A particularly important topic with factorial designs is the interaction between the factors, specifically, how to determine whether the effect of onw factor depends on the levels of other factors in the ANOVA.

Appendix 3-1: Summary of ANOVA Variations for Contrasts

```
ONEWAY psypath BY major /CONTRAST = −1 −1 −1 3.


GLM psypath BY major /CONTRAST(major) = SPECIAL(−1 −1 −1 3).


GLM psypath BY major
      /CONTRAST(major) = SPECIAL(−1 −1 −1 3  −1 −1 2 0  −1 1 0 0).


GLM psypath BY major
      /CONTRAST(major) = SPECIAL(−1 −1 −1  3)
      /CONTRAST(major) = SPECIAL(−1 −1  2  0)
      /CONTRAST(major) = SPECIAL(−1  1  0  0).


GLM psypath BY major
      /LMATRIX major −1 −1 −1 3.


MANOVA psypath BY major(1 4)
      /CONTRAST(major) = SPECIAL(1 1 1 1  −1 −1 −1 3   −1 −1 2 0  −1 1 0 0).


MANOVA psypath BY major(1 4) /PRINT = SIGNIFICANCE(SINGLEDF)
      /CONTRAST(major) = SPECIAL(1 1 1 1  −1 −1 −1 3   −1 −1 2 0  −1 1 0 0).


MANOVA psypath BY major(1 4)
      /CONTRAST(major) = SPECIAL(1 1 1 1  −1 −1 −1 3   −1 −1 2 0  −1 1 0 0)
      /DESIGN major(1) major(2) major(3).
```

APPENDIX 3-2: PRACTICE PROBLEMS FOR PLANNED COMPARISONS

For the following studies, generate a contrast with a theoretical rationale. Then, generate additional orthogonal contrasts (with or without a rationale) until you have k - 1 orthogonal contrasts. Assuming $n_j = 10$ observations in each condition, calculate $SS_{Treatment}$ and show the partitioning of $SS_{Treatment}$ into SS_Ls as shown for example 1. Generate predictions based on the conditions and assuming you did not see the results.

1. Educational psychologists examined student free recall after using different study methods: 1. No note-taking, review lecturer's notes. 2. No note-taking, mental review. 3. Note-taking, mental review. 4. Note-taking, review own notes. 5. Note-taking, review lecturer's notes.

	1	2	3	4	5	L	SS_L
$\bar{y}_j$	19.2	12.4	15.3	21.8	17.8		
Con_1	-3	-3	+2	+2	+2	15.0	75.0
Con_2	-1	+1	0	0	0	-6.8	231.2
Con_3	0	0	-2	1	1	9.0	135.0
Con_4	0	0	0	-1	1	-4.0	80.0
						Sum SS_L	521.2 = $SS_{Treatment}$

2. To test the efficacy of systematic desensitization for treatment of public speaking anxiety, Paul measured anxiety following subject participation in one of the following conditions.

	1. Control	2. Placebo	3. Insight Therapy	4. Systematic Desensitization
$\bar{y}_j$	36.0	33.4	32.8	24.6

3. Aggression was measured in children after exposure to one of the following five conditions.

	1. Aggressive Model	2. Filmed Agg Mod	3. Cartoon Agg Mod	4. No Model	5. Non-aggressive Model
$\bar{y}_j$	80	90	95	50	40

4. Infants were tested for depth perception using the visual cliff. Mothers encouraged their infants to cross from the shallow to the "deep" side with varying heights between the glass surface and the textured pattern below. Heights were set at 0, 10, 20, 40, and 80 cm.

5. Bandura's studies of modelling aggression towards a Bobo doll, included the following conditions. Analyze Males only.

	Model->	Aggressive		Non-Aggressive		Control
		F	M	F	M	
Female		5.5	7.2	2.5	0.0	1.2
Male	$\bar{y}_j$	12.4	25.8	0.2	1.5	2.0

6. Frequency of fighting was measured in female rats following injection of:

	Placebo	Estradiol	Testosterone
$\bar{y}_j$	1.8	2.1	5.0

7. Aggression in female gorillas was observed when the females were present with:

	Young	Females	Dominant Males	Adolescent Males
$\bar{y}_j$	20	18	5	10

8. Minutes of play was observed by blanket-attached (BA) and non-blanket-attached (NBA) children when they were alone in a novel environment with only the following object. Analyze results only for BA children.

		Mother	Blanket	Hard Toy	Nothing
	NBA	115	40	35	25
$\bar{y}_j$	BA	110	105	50	20

9. English and Hindi-speaking adults and infants were tested on their ability to discriminate two Hindi "t" sounds (same in English).

	Hindi Adult	English 7 mth	English 9 mth	English 11 mth	English Adult
$\bar{y}_j$	100	90	70	20	10

10. Time spent doing housework was observed in:

	Single Women	Single Men	Married Women	Married Men
$\bar{y}_j$	10	5	20	15

11. Amount of aggression between pairs of children was observed in the playground between pairs of Boys (B) and Girls (G).

	B on B	B on G	G on B	G on G
$\bar{y}_j$	40	10	11	8

12. Aggression was observed in adults classified by their scores on a measure of masculinity and feminity.

	Masculine	Feminine	Androgynous	Undifferentiated
$\bar{y}_j$	40	5	30	30

13. Frequency of discrimination was measured in the following four groups of Canadians.

	White	Chinese	South Asian	Black
$\bar{y}_j$	5	15	20	30

14. Participants who were Japanese or Caucasian were asked to describe couples that were either Japanese or Caucasian. Researchers measured the frequency of use of ethnicity-related terms in the descriptions.

Participant –>	Japanese	Japanese	Caucasian	Caucasian
Couple –>	Japanese	Caucasian	Japanese	Caucasian
$\bar{y}_j$	11	24	32	13

15. Americans from different ethnic groups were asked to rate how important their ethnic identity was to them.

	White	African	Asian	Hispanic
$\bar{y}_j$	4.59	5.73	5.29	5.45

16. Single mothers were observed interacting with their sons. The frequency of imperatives (orders) was measured.

	Middle		Working Class	
	Class	Skilled	Unskilled Dad+	Unskilled Dad-
$\bar{y}_j$	15	50	45	47

17. Number of hours per week that teenagers spend with their peers was measured.

	America	Canada	Japan	Taiwan	Beijing
$\bar{y}_j$	18	15	12	8	9

18. The frequency with which friends of grade 4 children make fun of people who try to do well in school was measured.

	White	African	Hispanic	Other
$\bar{y}_j$	18	36	30	25

19. Negative attitudes (do not accept) toward homosexuality were measured in people from different religious groups.

	Fundamentalist	Protestant	Catholic	Jewish	No Religion
$\bar{y}_j$	75	40	35	4	11

20. Positive attitudes toward mental illness were measured.

	Non-Greek Canadian	Greek Canadian	Greek Resident
$\bar{y}_j$	5.28	4.51	4.25

21. Number of words remembered was measured after Incidental or Intentional learning, with Incidental participants performing shallow (CheckEs, #Letters) or deep (Meaning) orienting tasks when studying the words.

	Incidental	Incidental	Incidental	Intentional
Orienting Task	Check Es	# Letters	Meaning	
$\bar{y}_j$	9	10	15	16

22. Memory for pairs of words was tested after words were learned under the following conditions.

	Repetition	Separate Imagery	Integrated Imagery
$\bar{y}_j$	35	45	80

23. Amount of organization used in learning by children who are developmentally slow or normal was tested at two age levels.

	Normal		Slow	
	6	8	6	8
$\bar{y}_j$	45	90	24	25

24. Memory for words was tested after participants Studied words under water or on land and were Tested under water or on land.

Studied –>	Water	Land	Water	Land
Tested –>	Land	Water	Water	Land
$\overline{y}_j$	20	19	34	37

25. Memory for an initial list of words was retested after interpolated learning of materials that varied in similarity to the original list.

	None	Numbers	Nonsense Syllables	Unrelated Words	Antonyms	Synonyms
$\overline{y}_j$	4.50	3.68	2.58	2.57	1.83	1.25

26. Subjects heard a lecture about the pros and cons of wearing seatbelts. Memory for pro and con statements was measured. Analyze results separately for Pro and Con statements.

	Participants Wear Seatbelt					
	Never		Sometimes		Always	
	Pro	Con	Pro	Con	Pro	Con
$\overline{y}_j$	1.60	2.07	1.72	1.78	2.29	1.61

CHAPTER 4 - BETWEEN-S FACTORIAL ANOVA

The single-factor Between-S design is appropriate when there is just one independent variable or factor in the study. Researchers often examine the effect of two or more factors at the same time. Such Between-S studies usually involve factorial designs in which subjects are randomly assigned or otherwise belong to all combinations of the factors. For example, a social psychology study involving attitude change about exercise might expose 60 subjects to attitude change messages in one of six conditions defined by two levels of Expertise of the source of information (Expert vs. Nonexpert) and three levels of Threat about lack of exercise (Low vs. Medium vs. High). The 60 subjects would be assigned equally to the 2 × 3 = 6 conditions (or cells) defined by the two factors, 10 subjects in each cell.

		Threat		
		Low	Med	High
Expertise	Expert	10	10	10
	Nonexpert	10	10	10

Factorial designs have several benefits. First, they are an efficient use of subjects. In this example, 10 subjects per cell would allow comparisons between Expert and Nonexpert (i.e., H_0: $\mu_{Expert}=\mu_{Nonexpert}$) to be based on 30 subjects in each condition averaged over the three levels of Threat. This is called a main effect, the effect of Expertise averaged across levels of Threat. Similarly, it would allow comparisons between Low, Medium, and High levels of threat (the main effect of Threat, H_0: $\mu_{Low}=\mu_{Medium}=\mu_{High}$) to be based on 20 subjects per cell averaged over the two levels of Expertise. Testing each of these effects in separate studies would require 2 × 60 = 120 subjects to have the same degree of power for both comparisons. The effect of one factor averaged over the levels of other factors in the analysis is called a "main effect."

Factorial designs also allow researchers to study interactions between factors. A statistical interaction means that the effect of one factor varies or differs across the levels of one or more other factors. For example, Threat might work differently for Expert and Nonexpert presentations. It might increase intention to exercise in the Expert condition but have no effect or even decrease intention to exercise in the Nonexpert condition. The effect of Threat would differ for the different levels of Expertise (Expert vs. Nonexpert). Interactions are often very important in psychological research, both for theoretical and practical reasons. Appendix 4-1 shows some real-life statistical interactions. Even if there was no interaction, the factorial design would demonstrate that the effect of each factor generalizes across the levels of the other factor. In this example, Threat might operate the same for both Expert and Nonexpert speakers.

Factorial designs vary with respect to whether each factor is Between-Subjects or Within-Subjects. The persuasion study as described above would involve two Between-S factors: 60 subjects randomly assigned with 10 subjects per cell. There would be no expectation that observations in one cell would correlate with observations in other cells. Less plausibly, the persuasion study could be done completely Within-S if just 10 subjects heard 6 different messages targeting a different behaviour and using all combinations of conditions. More plausibly, 60 subjects could be tested on some trait relevant to the dependent variable (e.g., gullibility?), formed into 10 blocks of 6 subjects matched on gullibility (i.e., everyone in a block had similar scores), and then assigned randomly one subject from each level of gullibility to each cell in the study. Despite different subjects in the six conditions, it is a Within-S design for analysis purposes because researchers would expect observations to be correlated across conditions. For example, the most gullible subject in each group would be most likely to change their mind in all conditions and the least gullible in each group would be least likely to change their mind. This study could also be done with one Within-S and one Between-S factor. Factorial designs can involve more than two factors. The persuasion study, for example, could include a third factor Gender, with half the participants Female and half Male. The study would now involve 2 × 3 × 2 = 12 cells or conditions. Chapters 4, 5, and 6 cover two-factor Between-S designs, and chapters 7, 8, and 9 cover analyses for designs that involve one or more Within-S factors.

A final observation. ANOVA is often associated with true experiments in which people are randomly assigned to conditions, but factors can be either experimental (e.g., Expertise and Threat) or non-experimental (e.g., Gender). The statistical analysis is the same although causal inferences are stronger for ***well-designed and executed*** experiments than for non-experimental studies. Poorly designed experiments are no better than non-experiments and sometimes even weaker sources for drawing valid causal inferences. Moreover, factorial analyses are simpler with equal numbers per cell, which could be less likely to occur in non-experimental studies unless participants were selected deliberately to ensure equal numbers.

Formulas for two-factor Between-S designs require a different notation than the single factor design. The basic notation for factorial ANOVA is described in Appendix 4-1. Briefly, capital letters A and B are names for the two factors ***and*** also the number of levels of each factor. Lowercase letters *a* and *b* index the different levels of A and B (comparable to j in the single factor design); that is, a = 1, 2, ..., A and b = 1, 2, ..., B for the levels of A and B, respectively. To illustrate, $\overline{y}_{ab} = \overline{y}_{23}$ denotes the mean for level 2 of factor A (a = 2) and level 3 of factor B (b = 3), and n_{14} is the sample size for level 1 of A and level 4 of B. If factor A had 3 levels and factor B had 4 levels, the number of cells would equal A×B = 3×4 = 12. With respect to main effects, $\overline{y}_a$ represents the mean for each level of A averaged over levels of B and $\overline{y}_b$ represents the mean for each level of B averaged over levels of A. And n_a and n_b would be the number of observations contributing to the main effect of A and B, respectively. For the entire set of observations, $\overline{y}_{Grand}$ is the mean of all scores and N is the total number of observations.

Calculations for Factorial ANOVA

The default factorial ANOVA partitions SS_{Total} into five components: SS_A is the main effect of A, SS_B is the main effect of B, and $SS_{A\times B}$ is the interaction between A and B. Error is any variability not accounted for by the main effects and the interaction. Hence,

$$SS_{Total} = SS_A + SS_B + SS_{A\times B} + SS_{Error}$$

The df for each SS is used to calculate four MSs, one being MS_{Error}, the denominator for all effects in the Between-S design. The remaining three MSs are each associated with a specific hypothesis about the means that can be tested by a corresponding F. MS_A and MS_B will test the main effects of factors A and B, respectively. These are called main effects because the means are averaged over the levels of the other factor. MS_{AxB} will test the significance of the interaction, namely whether the effect of factor A differs across levels of factor B, or vice versa.

The analysis is illustrated for a 2×4 factorial study of mistakes in a selective attention task with various distracting sounds. Sounds of two Types (factor A, random Noise or Speech) were played at four levels of Volume (factor B; Subthreshold, Audible, Normal Speech, and Shouting). A total of 24 subjects participated in the experiment, with 3 subjects assigned to each of the 2×4 = 8 cells of the experiment. The results appear in the following table with some calculations explained next.

					Volume (B)					
Type (A)		1. Subthr		2. Audible		3. Speech		4. Shout	$\bar{y}_a$	n_a
		1		4		5		3		
1. Noise		4		2		7		5		
		7		3		3		4		
$\bar{y}_{ab}$	$\bar{y}_{11}$	**4.0**	$\bar{y}_{12}$	**3.0**	$\bar{y}_{13}$	**5.0**	$\bar{y}_{14}$	**4.0**	**4.0**	**12**
n_{ab}	n_{11}	**3**	n_{12}	**3**	n_{13}	**3**	n_{14}	**3**		
		3		5		10		6		
2. Speech		2		2		9		10		
		1		8		8		8		
$\bar{y}_{ab}$	$\bar{y}_{21}$	**2.0**	$\bar{y}_{22}$	**5.0**	$\bar{y}_{23}$	**9.0**	$\bar{y}_{24}$	**8.0**	**6.0**	**12**
n_{ab}	n_{21}	**3**	n_{22}	**3**	n_{23}	**3**	n_{24}	**3**		
$\bar{y}_b$		**3.0**		**4.0**		**7.0**		**6.0**	$\bar{y}_G$**=5.0**	**N=24**
n_b		**6**		**6**		**6**		**6**	s_G**=2.798**	

SS Total. To calculate SS_{Total}, ignore all levels of both factors, and think of the dataset as 24 individual observations with $\bar{y}_G = 6.0$. We would subtract $\bar{y}_G$ from each of the 24 observations, square the deviations, and sum them up over all subjects within a group, and then over all levels of the Volume factor and all levels of the Type factor. This summing is represented by $\sum\sum\sum$ below. See Box 4.1 for the full notation.

$$
\begin{aligned}
SS_{Total} &= \sum\sum\sum(y_{abi} - \overline{y}_G)^2 = (y_{111} - \overline{y}_G)^2 + (y_{112} - \overline{y}_G)^2 + ... + (y_{242} - \overline{y}_G)^2 + (y_{243} - \overline{y}_G)^2 \\
&= (1\text{-}5.0)^2 + (4\text{-}5.0)^2 + (7\text{-}5.0)^2 + ... + (3\text{-}5.0)^2 + (5\text{-}5.0)^2 + (4\text{-}5.0)^2 \\
&\quad + (3\text{-}5.0)^2 + (2\text{-}5.0)^2 + (1\text{-}5.0)^2 + ... + (6\text{-}5.0)^2 + (10\text{-}5.0)^2 + (8\text{-}5.0)^2 \\
&= -5^2 + -2^2 + ... + 4^2 + 2^2 \\
&= 180.0 \\
df &= N - 1 = 24 - 1 = 23
\end{aligned}
$$

The $df = N - 1$ because one $\overline{y}_G$ was subtracted from N observations. If the overall standard deviation for the entire set of data is known, as above, then:

$$SS_{Total} = (N\text{-}1)s_G^2 = (24\text{-}1)2.798^2 = 180.062 \approx 180.0$$

This represents the total variability in the 24 scores, which can partitioned into error, main effect of Type, main effect of Volume, and the interaction between Type and Volume.

SS Error. As in the single factor Between-S design, error is the deviation of scores within each unique condition from the mean for that condition. In factorial designs, unique conditions are defined by a level of factor A (Type) and a level of factor B (Volume). Thus, there are 8 unique conditions in this study, each with its own cell mean, $\overline{y}_{ab}$. SS_{Error} is the deviations of the three observations in each condition from $\overline{y}_{ab}$, squared, summed across the three (n_{ab}) observations in each condition, and then across the four levels of B and the two levels of A (i.e., over all 8 conditions). These operations are illustrated below. Note as well that SS_{Error} can be obtained by summing SS_{ab} across all conditions and SS_{ab} for each condition can be obtained from s_{ab} if known.

$$
\begin{aligned}
SS_{Error} &= \sum\sum\sum(y_{abi} - \overline{y}_{ab})^2 = (y_{111} - \overline{y}_{11})^2 + (y_{112} - \overline{y}_{11})^2 + ... + (y_{242} - \overline{y}_{24})^2 + (y_{243} - \overline{y}_{24})^2 \\
&= (1\text{-}4.0)^2 + (4\text{-}4.0)^2 + (7\text{-}4.0)^2 + ... + (3\text{-}4.0)^2 + (5\text{-}4.0)^2 + (4\text{-}4.0)^2 \\
&\qquad + (3\text{-}2.0)^2 + (2\text{-}2.0)^2 + (1\text{-}2.0)^2 + ... + (6\text{-}8.0)^2 + (10\text{-}8.0)^2 + (8\text{-}8.0)^2 \\
&= -3^2 + 0^2 + 3^2 ...+ -2^2 + 2^2 + 0^2 \\
&= 60.0 \\
&= \sum\sum SS_{ab} = SS_{11} + SS_{12} + ... + SS_{24} = 18.0 + ... + 8.0 \\
&= \sum\sum(n_{ab} - 1)s_{ab}^2 = (n_{11}\text{-}1)s^2_{11} + (n_{12}\text{-}1)s^2_{12} + ... + (n_{24}\text{-}1)s^2_{24} = (3\text{-}1)\times 3.0^2 + ...(3\text{-}1)\times 2.0^2 \\
df &= \sum\sum(n_{ab} - 1) = (n_{11} - 1) + (n_{12} - 1) + ... + (n_{24} - 1) = (3\text{-}1) + ... \\
&= 16 \\
&= N - A\times B = 24 - 2\times 4 = 24 - 8 = 16
\end{aligned}
$$

SS Main Effects. To calculate SS for the main effects of A and B, treat the study as though the other factor does not exist. Averaged across the Volume factor, Type (factor A) produces two means, each based on 12 observations (n_a); these are denoted in general as $\overline{y}_a$. For the first level of A (i.e., a = 1), $\overline{y}_{1.} = 4.0$, and for the second level of A (i.e., a = 2), $\overline{y}_{2.} = 6.0$. The period (.) in the subscript indicates that 1 and 2 refer to the levels for A, the first factor, rather than B. Similarly, averaged across the Type factor, Volume (or B) produces four means, each based on 6 observations (n_b); these are denoted in general as $\overline{y}_b$, giving $\overline{y}_{.1} = 3.0$, $\overline{y}_{.2} = 4.0$, $\overline{y}_{.3} = 7.0$, and $\overline{y}_{.4} = 6.0$. The period in the first position indicates that the subscripts refer to levels of

B, the second factor. SS_A and SS_B can now be calculated essentially in the same manner as the single-factor SS for treatment. That is,

$$SS_A = \sum n_a(\bar{y}_a - \bar{y}_G)^2 = 12(4.0 - 5.0)^2 + 12(6.0 - 5.0)^2$$
$$= 12\times -1.0^2 + 12\times 1.0^2 \quad \text{or } 12\times(-1^2 + 1^2) \text{ when } n_a\text{s are equal}$$
$$= 24.0$$
$$df_A = A - 1 = 2 - 1 = 1$$

$$SS_B = \sum n_b(\bar{y}_b - \bar{y}_G)^2 = 6(3.0 - 5.0)^2 + 6(4.0 - 5.0)^2 + 6(7.0 - 5.0)^2 + 6(6.0 - 5.0)^2$$
$$= 6(-2.0^2 + -1.0^2 + 2.0^2 + 1.0^2)$$
$$= 60.0$$
$$df_B = B - 1 = 4 - 1 = 3$$

SS Interaction. We later consider ways to compute $SS_{A\times B}$ directly from sample means, but it can also be calculated by subtraction. $SS_{A\times B}$ is variability left over from SS_{Total} after SS_{Error}, SS_A and SS_B are removed; that is,

$$SS_{A\times B} = SS_{Total} - SS_{Error} - SS_A - SS_B = 180.0 - 60.0 - 24.0 - 60.0 = 36.0$$
$$df_{A\times B} = df_{Total} - df_{Error} - df_A - df_B = 23 - 16 - 1 - 3 = 3 = (A-1)(B-1) = (2-1) \times (4-1) = 1 \times 3$$

These calculations are summarized in Box 4-1 and produce all quantities needed for the Between-S two-factor ANOVA, shown below. The results correspond exactly to SPSS analyses that follow. The H_0s for Type and Volume are rejected, analogous to H_0 in the single factor design. The H_0 for T×V is No Interaction, explained more fully later. Here the observed F for the interaction is close to significance, sometimes referred to as "marginally significant" in journal articles.

$$SS_{Total} = \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{s=1}^{n_{ab}} (y_{abs} - \bar{y}_G)^2$$
$$SS_{Error} = \sum\sum\sum (y_{abs} - \bar{y}_{ab})^2$$
$$SS_A = \sum_{a=1}^{A} n_a(\bar{y}_a - \bar{y}_G)^2$$
$$SS_B = \sum_{b=1}^{B} n_b(\bar{y}_b - \bar{y}_G)^2$$
$$SS_{A\times B} = SS_{Total} - SS_A - SS_B - SS_{Error}$$

Box 4-1. ANOVA Formula

Source	*SS*	*df*	*MS*	*F*	*df*	$F_{.05}$	
Type	24.0	1	24.00	6.40	1, 16	4.49	Rej H_0: $\mu_N = \mu_S$.
Volume	60.0	3	20.00	5.33	3, 16	3.24	Rej H_0: $\mu_{V1} = \mu_{V2} = \mu_{V3} = \mu_{V4}$
T x V	36.0	3	12.00	3.20	3, 16	3.24	"Do Not Rej" H_0: ??
Error	60.0	16	3.75				
Total	180.0	23					

SPSS Analyses for Between-S Factorial Design

The standard way to enter data for two-factor Between-S designs is to enter three values for each case: one for the level of factor A (*typ*), a second for the level of factor B (*vol*), and a third for the dependent variable, mistakes (*mis*) in this study. The following syntax creates a data set with 24 rows and 3 values for each row or case: level for Type, level for Volume, and number of mistakes.

The GLM command allows for multiple BS factors after the BY term, here *typ* and *vol*, and the PLOT option graphs main effects or interactions. The descriptive statistics provide enough information to complete the ANOVA calculations. Operations performed earlier to calculate SSs manually are shown to the right of the descriptive statistics.

```
DATA LIST FREE / typ vol mis.
BEGIN DATA
1 1  1      1 1  4      1 1  7      1 2  4      1 2  2      1 2  3
1 3  5      1 3  7      1 3  3      1 4  3      1 4  5      1 4  4
2 1  3      2 1  2      2 1  1      2 2  5      2 2  2      2 2  8
2 3 10      2 3  9      2 3  8      2 4  6      2 4 10      2 4  8
END DATA.

GLM mis BY typ vol /PRINT = DESCR /PLOT = PROFILE(vol BY typ).
```

typ	vol	Mean	Std. Deviation	N	SS_{Total}	SS_{Error}	$n_a(\bar{y}_a - \bar{y}_G)^2$	$n_b(\bar{y}_b - \bar{y}_G)^2$
1	1	4.00	3.000	3		$(3-1)3.0^2$		
	2	3.00	1.000	3		$(3-1)1.0^2$		
	3	5.00	2.000	3		...		
	4	4.00	1.000	3		...		
	Total	4.00	1.809	12			$12(4.0-5.0)^2$	
2	1	2.00	1.000	3		...		
	2	5.00	3.000	3		...		
	3	9.00	1.000	3		...		
	4	8.00	2.000	3		...		
	Total	6.00	3.303	12			$12(6.0-5.0)^2$	
Total	1	3.00	2.280	6				$6(3.0-5.0)^2$
	2	4.00	2.280	6				$6(4.0-5.0)^2$
	3	7.00	2.608	6				$6(7.0-5.0)^2$
	4	6.00	2.608	6				$6(6.0-5.0)^2$
	Total	5.00	2.798	24	$(24-1)2.798^2$			

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Corrected Model	120.000(a)	7	17.143	4.571	.006	***= $SS_A + SS_B + SS_{AxB}$***
Intercept	600.000	1	600.000	160.000	.000	***= $24 \times (5.0 - 0)^2$***
typ	**24.000**	**1**	**24.000**	**6.400**	**.022**	
vol	**60.000**	**3**	**20.000**	**5.333**	**.010**	
typ * vol	**36.000**	**3**	**12.000**	**3.200**	**.052**	
Error	**60.000**	**16**	**3.750**			
Total	780.000	24				***$600.0 + SS_{Total}$***
Corrected Total	**180.000**	**23**				***SS_{Total}***

a R Squared = .667 (Adjusted R Squared = .521) ***$R^2 = 120.0/180.0$***

The GLM results in bold agree with earlier calculations. Dividing *SS* by *df* produces the four MSs required to test the three hypotheses. MS_{Error} = 60.0 / 16 = 3.75 is the denominator for the three critical tests. The notes above explain some of the secondary quantities calculated by GLM. Note in particular that our SS_{Total} is the Corrected Total in GLM. The significance values for the main effect tests lead to the following conclusions, identical to the earlier conclusions based on critical values.

Reject H_0: $\mu_{Noise} = \mu_{Speech}$ *Sig. = .022*

Reject H_0: $\mu_{V1} = \mu_{V2} = \mu_{V3} = \mu_{V4}$ *Sig. = .010*

In both cases, we reject H_0 and accept H_a that one or more equalities is false. The conclusion about the interaction is somewhat ambiguous; as shown earlier, the *F* is very close to significant, $p = .052$, and the graph in Figure 4-1 indicates an interaction. Specifically, the effect of Volume appears different for the Noise and Speech distractors, being much stronger for Speech. Equivalently, the difference between Speech and Noise is larger for high levels of Volume (i.e., 3 and 4) than for low levels of Volume (1 and 2). Later examination of interaction analyses will show that the standard *F* test for A×B is optimal only for complete cross-over interactions and can be insensitive to many observed interactions. That is, the default test is most sensitive when all variability between conditions is due to interaction and none to main effects.

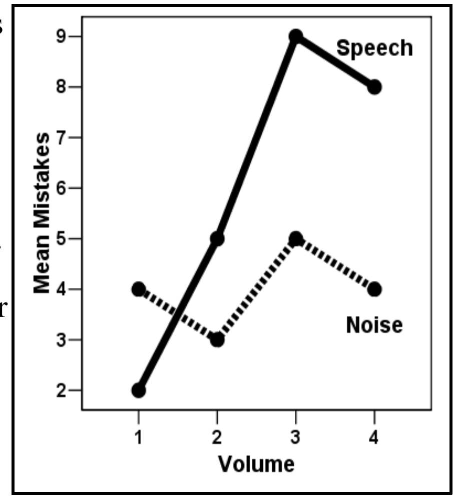


Figure 4-1. Cell Means.

The following MANOVA analysis duplicates the various quantities calculated earlier and produced by GLM.

```
MANOVA mis BY typ(1 2) vol(1 4) /PRINT = CELL.
      FACTOR            CODE        Mean  Std. Dev.          N
  typ                     1
   vol                    1        4.000      3.000          3
   vol                    2        3.000      1.000          3
   vol                    3        5.000      2.000          3
   vol                    4        4.000      1.000          3
  typ                     2
   vol                    1        2.000      1.000          3
   vol                    2        5.000      3.000          3
   vol                    3        9.000      1.000          3
   vol                    4        8.000      2.000          3

 For entire sample                 5.000      2.798         24

 Source of Variation          SS       DF         MS         F  Sig of F
 WITHIN CELLS              60.00       16       3.75
 typ                       24.00        1      24.00      6.40      .022
 vol                       60.00        3      20.00      5.33      .010
 typ BY vol                36.00        3      12.00      3.20      .052

 (Model)                  120.00        7      17.14      4.57      .006
 (Total)                  180.00       23       7.83
```

Using SPSS to Compute SSs

There are several ways for SPSS to calculate SSs for the Between-S factorial. One method uses GLM to compute and save predicted (and residual) values. The next few analyses illustrate the process. The various scores produced are shown after the final analysis. ANOVA output is deleted.

GLM mis /SAVE PRED(mg).	***Saves*** $\bar{y}_G$ ***as mg***
...	
COMPUTE tot = mis – mg.	$y - \bar{y}_G$ ***= Residual if saved in GLM above***
COMPUTE tot2 = tot2.**	$(y - \bar{y}_G)^2$
DESCR tot2 /STAT = SUM.	
N Sum	
tot2 24 180.0000	$= SS_{Total}$
GLM mis BY typ vol /SAVE PRED(mab).	***Saves*** $\bar{y}_{ab}$ ***as mab***
COMPUTE err = mis – mab.	$y - \bar{y}_{ab}$ ***= Residual if saved in GLM above***
COMPUTE err2 = err2.**	
DESCR err2 /STAT = SUM.	
N Sum	
err2 24 60.0000	$= SS_{Error}$
GLM mis BY typ /SAVE PRED(ma).	***Saves*** $\bar{y}_a$ ***as ma***
COMPUTE amain = ma – mg.	$\bar{y}_a - \bar{y}_G$
COMPUTE amain2 = amain2.**	
DESCR amain2 /STAT = SUM.	
N Sum	
amain2 24 24.0000	$= SS_A = SS_{Type}$
GLM mis BY vol /SAVE PRED(mb).	***Saves*** $\bar{y}_b$ ***as mb***
COMPUTE bmain = mb – mg.	$\bar{y}_b - \bar{y}_G$
COMPUTE bmain2 = bmain2.**	
DESCR bmain2 /STAT = SUM.	
N Sum	
bmain2 24 60.0000	$= SS_B = SS_{Volume}$

The following table shows the variables created by the preceding commands. Squared deviations are not shown.

LIST typ vol mis mg tot mab err ma amain mb bmain.

		y	$\bar{y}_G$	$y-\bar{y}_G$	$\bar{y}_{ab}$	$y-\bar{y}_{ab}$	$\bar{y}_a$	$\bar{y}_a-\bar{y}_G$	$\bar{y}_b$	$\bar{y}_b-\bar{y}_G$
typ	**vol**	**mis**	**mg**	**tot**	**mab**	**err**	**ma**	**amain**	**mb**	**bmain**
1	1	1	5.0	-4.0	4.0	-3.0	4.0	-1.0	3.0	-2.0
1	1	4	5.0	-1.0	4.0	.0	4.0	-1.0	3.0	-2.0
1	1	7	5.0	2.0	4.0	3.0	4.0	-1.0	3.0	-2.0
1	2	4	5.0	-1.0	3.0	1.0	4.0	-1.0	4.0	-1.0
1	2	2	5.0	-3.0	3.0	-1.0	4.0	-1.0	4.0	-1.0
1	2	3	5.0	-2.0	3.0	.0	4.0	-1.0	4.0	-1.0
1	3	5	5.0	.0	5.0	.0	4.0	-1.0	7.0	2.0
1	3	7	5.0	2.0	5.0	2.0	4.0	-1.0	7.0	2.0
1	3	3	5.0	-2.0	5.0	-2.0	4.0	-1.0	7.0	2.0
1	4	3	5.0	-2.0	4.0	-1.0	4.0	-1.0	6.0	1.0
1	4	5	5.0	.0	4.0	1.0	4.0	-1.0	6.0	1.0
1	4	4	5.0	-1.0	4.0	.0	4.0	-1.0	6.0	1.0

2	1	3	5.0	-2.0	2.0	1.0	6.0	1.0	3.0	-2.0	
2	1	2	5.0	-3.0	2.0	.0	6.0	1.0	3.0	-2.0	
2	1	1	5.0	-4.0	2.0	-1.0	6.0	1.0	3.0	-2.0	
2	2	5	5.0	.0	5.0	.0	6.0	1.0	4.0	-1.0	
2	2	2	5.0	-3.0	5.0	-3.0	6.0	1.0	4.0	-1.0	
2	2	8	5.0	3.0	5.0	3.0	6.0	1.0	4.0	-1.0	
2	3	10	5.0	5.0	9.0	1.0	6.0	1.0	7.0	2.0	
2	3	9	5.0	4.0	9.0	.0	6.0	1.0	7.0	2.0	
2	3	8	5.0	3.0	9.0	-1.0	6.0	1.0	7.0	2.0	
2	4	6	5.0	1.0	8.0	-2.0	6.0	1.0	6.0	1.0	
2	4	10	5.0	5.0	8.0	2.0	6.0	1.0	6.0	1.0	
2	4	8	5.0	3.0	8.0	.0	6.0	1.0	6.0	1.0	
				$\sum^2 = SS_{Total}$		**$\sum^2 = SS_{Error}$**		**$\sum^2 = SS_{Type}$**		**$\sum^2 = SS_{Volume}$**	

Computing SS for the Interaction

Subtracting SSs to obtain the interaction is correct and shows that the interaction is variability not accounted for by main effects or error, but it is inadequate for understanding what the actual quantity $SS_{A\times B}$ represents. Calculating the interaction directly reveals more clearly what aspect of the data is captured by $SS_{A\times B}$ and also explicitly shows the partitioning of SS_{Total}. Analogous to subtracting SSs from SS_{Total}, $SS_{A\times B}$ can be calculated by determining if there is any variability left in the cell means (the $\bar{y}_{ab}$s) when main effects are removed. Or we could determine whether the observed cell means differ from the cell means expected if there were only main effects and no interaction. Both approaches lead to the same conclusion and equal $SS_{A\times B}$ as calculated by subtraction.

Below are the 8 cell means ($\bar{y}_{ab}$) calculated earlier, along with the row means ($\bar{y}_a$), the column means ($\bar{y}_b$), and the grand mean ($\bar{y}_G$). The main effect of A (i.e., deviation from $\bar{y}_G$) is shown in the column headed $\bar{y}_a$ - $\bar{y}_G$, and the main effect of B is shown in the row labelled $\bar{y}_b$ - $\bar{y}_G$. Subtracting these effects from the cell means removes main effects (see formula in Box 4-2). The subsequent two tables show the results of subtracting these main effects.

$$SS_{AxB} = n_{ab} \times \sum_{a=1}^{A} \sum_{b=1}^{B} [\{\bar{y}_{ab} - (\bar{y}_a - \bar{y}_G) - (\bar{y}_b - \bar{y}_G)\} - \bar{y}_G]^2$$

Box 4-2. Cell Means Minus Main Effects.

		Volume (B)					
	$\bar{y}_{ab}$	1	2	3	4	$\bar{y}_a$	$\bar{y}_a$ - $\bar{y}_G$
Type (A)	1	4	3	5	4	4.0	-1.0
	2	2	5	9	8	6.0	+1.0
	$\bar{y}_b$	3.0	4.0	7.0	6.0	$\bar{y}_G$ = 5.0	
	$\bar{y}_b - \bar{y}_G$	-2.0	-1.0	+2.0	+1.0		

$\bar{y}'_{ab} = \bar{y}_{ab} - (\bar{y}_b - \bar{y}_G)$

	1	2	3	4	$\bar{y}_a'$	$\bar{y}_a' - \bar{y}_G$
1	6	4	3	3	4.0	-1.0
2	4	6	7	7	6.0	+1.0
$\bar{y}'_b$	5.0	5.0	5.0	5.0	$\bar{y}_G$ 5.0	

$\bar{y}''_{ab} = \bar{y}'_{ab} - (\bar{y}_a' - \bar{y}_G)$

	1	2	3	4	$\bar{y}_a''$
1	7	5	4	4	5.0
2	3	5	6	6	5.0
$\bar{y}''_b$	5.0	5.0	5.0	5.0	$\bar{y}_G$ 5.0

The bottom left table shows the result of subtracting the main effect of B. The two original means for Volume 1 are adjusted by subtracting the B effect of Volume 1, which is -2.0; to illustrate, $\overline{y}_{11}$ = 4.0, and $\overline{y}'_{11}$ = 4.0 - -2.0 = 6.0, and $\overline{y}_{21}$ = 2.0, and $\overline{y}'_{21}$ = 2.0 - -2.0 = 4.0. Cell means for Volumes 2, 3, and 4 are adjusted by subtracting their Volume main effects: -1.0, +2.0, and +1.0, respectively. The adjusted cell means are denoted $\overline{y}'_{ab}$. The main effect means for factor B using these adjusted cell means all equal the grand mean; that is, all $\overline{y}'_{b}$s = 5.0 = $\overline{y}_{G}$. The main effect of B has been removed or eliminated.

The bottom right table removes the main effects for A from each $\overline{y}'_{ab}$ in a similar manner, producing a $\overline{y}''_{ab}$ for each cell free of both main effects. For example, $\overline{y}''_{11}$ = 6.0 - -1.0 = 7.0. Now all the row and column main effect means equal $\overline{y}_{G}$ = 5.0. If there was only main effects in the results and no interaction, the 8 adjusted cell means $\overline{y}''_{ab}$ would all also equal $\overline{y}_{G}$ = 5.0. The fact that the adjusted cell means differ from 5.0 indicates the presence of an interaction in the data. That is, there is variability due to the unique combination of specific levels of Factor A and specific levels of Factor B, over and above the main effects of A and B, which have been removed.

The final step to calculate $SS_{Interaction}$ is: subtract $\overline{y}_{G}$ from each $\overline{y}''_{ab}$, square the deviations, multiply by the number of observations in each cell (i.e., n_{ab} = 3), and sum to get a total. The resulting deviations are shown in the next table. Therefore, $SS_{Interaction} = 3\times(+2^2 + 0^2 + -1^2 + -1^2 + -2^2 + 0^2 + 1^2 + 1^2) = 36.0$, the value obtained earlier by subtracting SSs. What these deviations mean is explained shortly.

$\overline{y}''_{ab}-\overline{y}_{G}$

Type	1	2	3	4
1	+2.0	0.0	-1.0	-1.0
2	-2.0	0.0	+1.0	+1.0

The second approach to $SS_{Interaction}$ is to calculate predicted cell means if there were only main effects and no interaction, and then determine how much the observed cell means ($\overline{\mathbf{y}}_{ab}$) deviate from the no interaction cell means (see formula in Box 4-3). These deviations will equal those just calculated; squaring, multiplying by n_{ab}, and summing will produce $SS_{Interaction}$. What this approach demonstrates is that $SS_{A\times B}$ represents how much the observed cell means deviate from those expected if there was no interaction.

$$SS_{A\times B}=n_{ab}\times\sum_{a=1}^{A}\sum_{b=1}^{B}[\overline{y}_{ab}-\{\overline{y}_{G}+(\overline{y}_{a}-\overline{y}_{G})+(\overline{y}_{b}-\overline{y}_{G})\}]^2$$

Box 4-3. Grand Mean Plus Main Effects.

To determine the expected cell means given no interaction, add the main effects of A and B to the grand mean, as illustrated in the left table below after the observed cell means. For example, $\overline{y}'_{11}$ = 5.0 + -1.0 + -2.0 = 2.0. The right table shows deviations of observed cell means ($\overline{y}_{ab}$ - $\overline{y}'_{ab}$) from the predicted cell means given no interaction; that is, ($\overline{y}_{ab}$ - $\overline{y}'_{ab}$). These deviations equal those calculated earlier.

$\bar{y}_{ab}$	Volume (B) 1	2	3	4	$\bar{y}_a$	$\bar{y}_a - \bar{y}_G$
Type (A) 1	4	3	5	4	4.0	−1.0
2	2	5	9	8	6.0	+1.0
$\bar{y}_b$	3.0	4.0	7.0	6.0	$\bar{y}_G$ = 5.0	
$\bar{y}_b - \bar{y}_G$	−2.0	−1.0	+2.0	+1.0		

$\bar{y}'_{ab} = \bar{y}_G + (\bar{y}_a - \bar{y}_G) + (\bar{y}_b - \bar{y}_G)$

	1	*2*	*3*	*4*	$\bar{y}_a$	$\bar{y}_a - \bar{y}_G$
1	2	3	6	5	4.0	−1.0
2	4	5	8	7	6.0	+1.0
$\bar{y}_b$	3.0	4.0	7.0	6.0	$\bar{y}_G$ 5.0	
$\bar{y}_b - \bar{y}_G$	−2.0	−1.0	+2.0	+1.0		

$\bar{y}_{ab} - \bar{y}'_{ab}$

	1	*2*	*3*	*4*
1	+2.0	0.0	−1.0	−1.0
2	−2.0	0.0	+1.0	+1.0

The critical aspect of the predicted cell means in the table on the left is that the main effects of A are the same at every level of B. The difference between Noise and Speech is exactly 2.0 units for Volumes 1, 2, 3, and 4, indicating 0 interaction. Equivalently, the main effect deviations for the two levels of A (i.e., -1 and +1) are the same at every level of Volume as shown by the deviation of cell means from column means. The main effects of B are also the same at every level of A since the main effect deviations for Volume (-2, -1, +2, +1) are identical for the Noise and Speech conditions, as shown by deviations of cell means from row means. That is, the deviations of 2.0, 3.0, 6.0, and 5.0 from the Noise row mean of 4.0 are -2, -1, +2, and +1, and the deviations of 4.0, 5.0, 8.0, and 7.0 from the Speech row mean of 6.0 are -2, -1, +2, and +1, equal to deviations for the Noise condition and the main effect of Volume.

A graph of $\bar{y}'_{ab}$ above would produce parallel lines, as shown shortly, indicating no interaction. The interaction deviations in the right table and the earlier table indicate how far away the observed cell means, $\bar{y}_{ab}$, are from parallel lines (i.e., from 0 interaction). The null hypothesis for the interaction would be that all these deviations would be 0 in terms of μ_{ab}, that is, deviations representing interaction would all be 0 in the population.

SPSS Calculation of $SS_{Interaction}$

Earlier analyses did much of the work necessary for GLM to compute $SS_{Interaction}$ when we computed the main effects and error using GLM. Specifically, GLM saved predicted scores for $\bar{y}_G$, $\bar{y}_{ab}$, $\bar{y}_a$, and $\bar{y}_b$ and then calculated $\bar{y}_a - \bar{y}_G$ (called *amain*) and $\bar{y}_b - \bar{y}_G$ (called *bmain*). These quantities are enough to compute the interaction according to the two procedures described above, first removing main effects.

```
COMPUTE mabsubmain = mab - amain - bmain.
COMPUTE intone = mabsubmain - mg.
COMPUTE intone2 = intone**2.
```

$\bar{y}_{ab} - (\bar{y}_a - \bar{y}_G) - (\bar{y}_b - \bar{y}_G)$

$\{\bar{y}_{ab} - (\bar{y}_a - \bar{y}_G) - (\bar{y}_b - \bar{y}_G)\} - \bar{y}_G$

```
DESCR intone2 /STAT = SUM.
                N  Sum
 intone2        24 36.0000
```

The second method is to generate predicted cell means for only main effects and subtract these values from the cell means. Two methods can be used to obtain predicted cell means with only main effects, the first using COMPUTE statements.

```
COMPUTE mabmain = mg + amain + bmain.
COMPUTE inttwo = mab - mabmain.
COMPUTE inttwo2 = inttwo**2.
```

$\bar{y}_G + (\bar{y}_a - \bar{y}_G) + (\bar{y}_b - \bar{y}_G)$

$\bar{y}_{ab} - \{\bar{y}_G + (\bar{y}_a - \bar{y}_G) + (\bar{y}_b - \bar{y}_G)\}$

```
DESCR inttwo2 /STAT = SUM.
```

	N	Sum
inttwo2	24	36.0000

Cell means for no interaction can also be obtained with the /DESIGN option on GLM. The default GLM tests main effects and interaction and corresponds to /DESIGN typ vol typ BY vol, whereas /DESIGN typ vol instructs GLM to calculate main effects only and use those main effects to generate predicted scores.

```
GLM mis BY typ vol /SAVE PRED(mabmaintwo) /DESIGN typ vol.
```

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	84.000(a)	4	21.000	4.156	.014
Intercept	600.000	1	600.000	118.750	.000
typ	24.000	1	24.000	4.750	.042
vol	60.000	3	20.000	3.958	.024
Error	96.000	19	5.053		
Total	780.000	24			
Corrected Total	180.000	23			

```
FORMAT typ vol (F1.0) mis (F2.0) mg TO mabmaintwo (F4.1).
LIST.
```

typ	vol	mis	mg	mab	ma	amain	mb	bmain	**mabsubmain**	**intone**	intone2	**mabmain**	**inttwo**	inttwo2	**mabmaintwo**
1	1	1	5.0	4.0	4.0	-1.0	3.0	-2.0	7.0	2.0	4.0	2.0	2.0	4.0	2.0
1	1	4	5.0	4.0	4.0	-1.0	3.0	-2.0	7.0	2.0	4.0	2.0	2.0	4.0	2.0
1	1	7	5.0	4.0	4.0	-1.0	3.0	-2.0	7.0	2.0	4.0	2.0	2.0	4.0	2.0
1	2	4	5.0	3.0	4.0	-1.0	4.0	-1.0	5.0	.0	.0	3.0	.0	.0	3.0
1	2	2	5.0	3.0	4.0	-1.0	4.0	-1.0	5.0	.0	.0	3.0	.0	.0	3.0
1	2	3	5.0	3.0	4.0	-1.0	4.0	-1.0	5.0	.0	.0	3.0	.0	.0	3.0
1	3	5	5.0	5.0	4.0	-1.0	7.0	2.0	4.0	-1.0	1.0	6.0	-1.0	1.0	6.0
1	3	7	5.0	5.0	4.0	-1.0	7.0	2.0	4.0	-1.0	1.0	6.0	-1.0	1.0	6.0
1	3	3	5.0	5.0	4.0	-1.0	7.0	2.0	4.0	-1.0	1.0	6.0	-1.0	1.0	6.0
1	4	3	5.0	4.0	4.0	-1.0	6.0	1.0	4.0	-1.0	1.0	5.0	-1.0	1.0	5.0
1	4	5	5.0	4.0	4.0	-1.0	6.0	1.0	4.0	-1.0	1.0	5.0	-1.0	1.0	5.0
1	4	4	5.0	4.0	4.0	-1.0	6.0	1.0	4.0	-1.0	1.0	5.0	-1.0	1.0	5.0
2	1	3	5.0	2.0	6.0	1.0	3.0	-2.0	3.0	-2.0	4.0	4.0	-2.0	4.0	4.0
2	1	2	5.0	2.0	6.0	1.0	3.0	-2.0	3.0	-2.0	4.0	4.0	-2.0	4.0	4.0
2	1	1	5.0	2.0	6.0	1.0	3.0	-2.0	3.0	-2.0	4.0	4.0	-2.0	4.0	4.0
2	2	5	5.0	5.0	6.0	1.0	4.0	-1.0	5.0	.0	.0	5.0	.0	.0	5.0
2	2	2	5.0	5.0	6.0	1.0	4.0	-1.0	5.0	.0	.0	5.0	.0	.0	5.0
2	2	8	5.0	5.0	6.0	1.0	4.0	-1.0	5.0	.0	.0	5.0	.0	.0	5.0
2	3	10	5.0	9.0	6.0	1.0	7.0	2.0	6.0	1.0	1.0	8.0	1.0	1.0	8.0
2	3	9	5.0	9.0	6.0	1.0	7.0	2.0	6.0	1.0	1.0	8.0	1.0	1.0	8.0
2	3	8	5.0	9.0	6.0	1.0	7.0	2.0	6.0	1.0	1.0	8.0	1.0	1.0	8.0
2	4	6	5.0	8.0	6.0	1.0	6.0	1.0	6.0	1.0	1.0	7.0	1.0	1.0	7.0
2	4	10	5.0	8.0	6.0	1.0	6.0	1.0	6.0	1.0	1.0	7.0	1.0	1.0	7.0
2	4	8	5.0	8.0	6.0	1.0	6.0	1.0	6.0	1.0	1.0	7.0	1.0	1.0	7.0
										$\sum^2 = SS_{Interaction}$			$\sum^2 = SS_{Interaction}$		

MANOVA and the Interaction Deviations

MANOVA can also generate the interaction deviations. The /PMEANS command in MANOVA generates predicted means based on the design. For the full factorial, predicted means equal observed cell means and there is no deviation or difference between predicted and observed means. If only main effects are specified for /DESIGN, predicted cell means are based on main effects and observed minus predicted means equal the interaction deviations. The next two analyses show the predicted means and residuals (deviations) for the full factorial and the main effect analyses. The /DESIGN typ vol typ BY vol in the first MANOVA is the default factorial and optional. Note Est. Mean and Raw Resid. columns differ in the two results.

```
MANOVA mis BY typ(1 2) vol(1 4) /PMEANS /DESIGN typ vol typ BY vol.
...
 Adjusted and Estimated Means
       Factor  Code   Obs. Mean  Adj. Mean     Est. Mean    Raw Resid.   Std. Resid.
  typ            1
   vol             1   4.00000    4.00000       4.00000        .00000        .00000
   vol             2   3.00000    3.00000       3.00000        .00000        .00000
   vol             3   5.00000    5.00000       5.00000        .00000        .00000
   vol             4   4.00000    4.00000       4.00000        .00000        .00000

  typ            2
   vol             1   2.00000    2.00000       2.00000        .00000        .00000
   vol             2   5.00000    5.00000       5.00000        .00000        .00000
   vol             3   9.00000    9.00000       9.00000        .00000        .00000
   vol             4   8.00000    8.00000       8.00000        .00000        .00000

MANOVA mis BY typ(1 2) vol(1 4) /PMEANS /DESIGN typ vol.
...
 Adjusted and Estimated Means
       Factor    Code  Obs. Mean   Adj. Mean     Est. Mean    Raw Resid.   Std. Resid.
  typ             1
   vol              1  4.00000     2.00000       2.00000       2.00000        .88976
   vol              2  3.00000     3.00000       3.00000        .00000        .00000
   vol              3  5.00000     6.00000       6.00000      -1.00000       -.44488
   vol              4  4.00000     5.00000       5.00000      -1.00000       -.44488

  typ             2
   vol              1  2.00000     4.00000       4.00000      -2.00000       -.88976
   vol              2  5.00000     5.00000       5.00000        .00000        .00000
   vol              3  9.00000     8.00000       8.00000       1.00000        .44488
   vol              4  8.00000     7.00000       7.00000       1.00000        .44488
```

In the second analysis, the Obs. Mean column shows the 8 cell means and the Adj. Mean or Est. Mean column shows means expected on the basis of main effects. These values equal those calculated previously. The Raw Resid. column shows the deviations of observed from expected means, which represent the interaction. Squaring these values and multiplying by $n_{ab}=3$ for each cell gives SS_{TxV}.

Graphing the Interaction and Related Quantities

Graphs are an excellent way to show interaction effects, and there are several methods in SPSS to graph the results of factorial ANOVAs. The basic factorial graph for a two-factor study uses the horizontal axis for one factor (usually the one with the most levels, here *vol* with four levels), and separate lines and markers for the other factors (here *typ* with two levels). The graph in Box 4-1 was created by the following GLM command, specifically the /PLOT = PROFILE(effects) option. The effects can be either main effects (e.g., PROFILE(vol)) or interactions, as shown here.

```
GLM mis BY typ vol /PLOT = PROFILE(vol BY typ).
```

This same graph could have been produced from menus by *Graph* | *Line* | *Multiple* | *Define*, which brings up the dialogue box in Figure 4-2. Users specify which factor goes on the Category Axis (*vol* has been inserted here) and which factor is used to Define Lines (*typ* here). Users also have the option of what to plot; here we want to plot the means for *mis*, our dependent variable. Clicking *Ok* creates the basic graph, which can be edited in the chart editor.

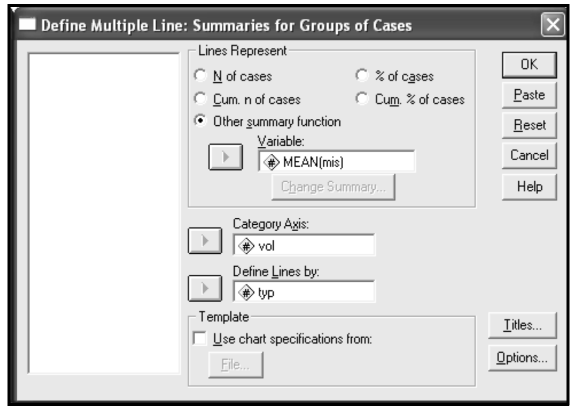


Figure 4-2. Graph menu

Understanding the factorial ANOVA, especially the interaction, benefits from plotting intermediate quantities produced previously, namely expected cell means with just main effects and no interaction, and adjusted cell means with main effects removed (i.e., pure interaction).

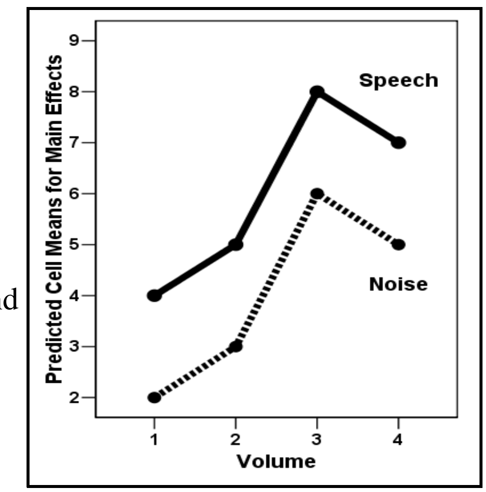


Figure 4-3. Predicted Cell Means if No Interaction

The original (Figure 4-1) and main effects only (Figure 4-3) graphs illustrate the logic of the interaction deviations. In essence the deviation for each cell represents how far its mean is from where it would be if there were only main effects and no interaction, as represented in Box 4-3. The observed mean for Speech at Volume 1, for example, is 2.0, which is 2 units lower than the 4.0 it would be if there was no interaction. The sum of the deviations squared times n_{ab} represents evidence for an interaction between Volume and Type of distractor in this study.

Figure 4-4 shows a plot of the adjusted cell means with main effects removed. This graph represents "pure" interaction. The mean score for each cell is 5.0 when averaged across levels of the other factor. In this graph the interaction is the deviation of the adjusted cell means from the grand mean of 5.0, which all cells would equal if there is no interaction.

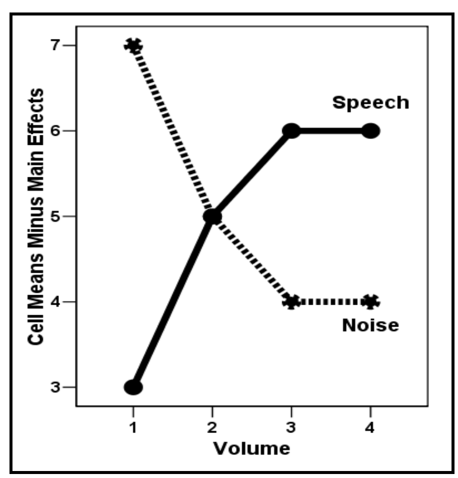


Figure 4-4. Cell Means Minus Main Effects

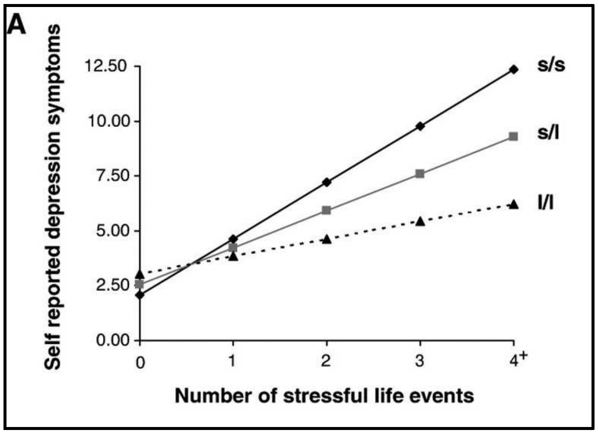
To illustrate, the adjusted cell mean for Speech at Volume 1 is 3.0, which is 2 units lower than the grand mean.

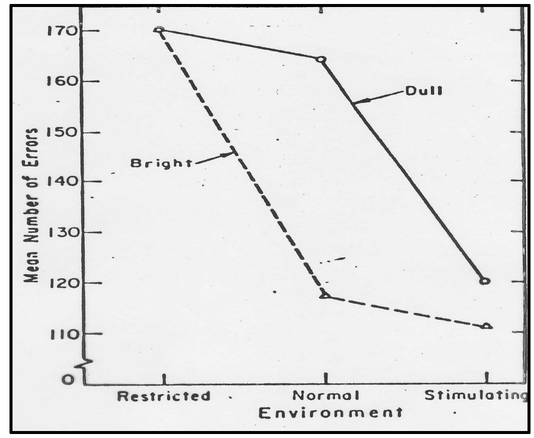
Although awkward, we could now state formally the H_0 for the interaction. Specifically, for all cells $\mu_{ab} = \mu_G + (\mu_a - \mu_G) + (\mu_b - \mu_G)$ or $\mu_{ab} - \{\mu_G + (\mu_a - \mu_G) + (\mu_b - \mu_G)\} = 0$. Much easier to say No Interaction.

The present study showed significant main effects for Volume and Type, and a marginally significant ($p = .052$) interaction. The results for a factorial design can produce any combination of significance for the three effects. One or both main effects can be significant or not, and the interaction can be significant or not, all independent of one another to a large degree depending on the nature of the interaction. Appendix 4-3 illustrates diverse outcomes that are possible for a factorial design with two factors. With more factors, more complex combinations of effects can be significant or not. And possible patterns for each factor become even more complex as the number of levels of each factor increase. Appendix 4-4 demonstrates the importance of sample size for statistical analyses to reach significance. In the distraction study, the marginally significant interaction becomes highly significant with more subjects.

As with single factor designs, rejecting the null hypothesis can be ambiguous and require further analyses of main and interaction effects. Or even if omnibus effects are not significant, some effects may warrant further investigation if certain patterns are predicted. These topics are examined next.

APPENDIX 4-1: EXAMPLES OF INTERACTIONS





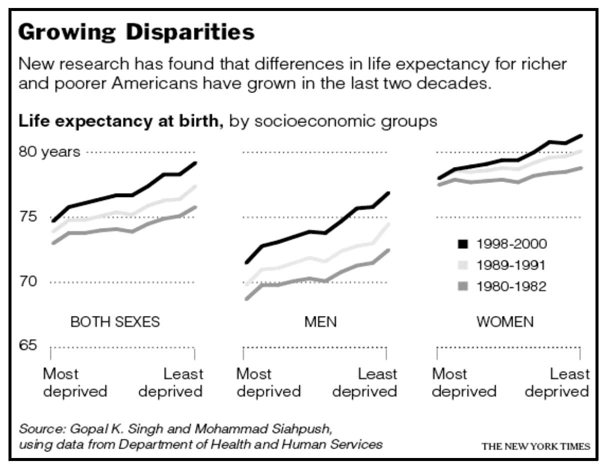
GxE Interaction: Maze learning errors for genetically bright and dull animals reared in different environments.

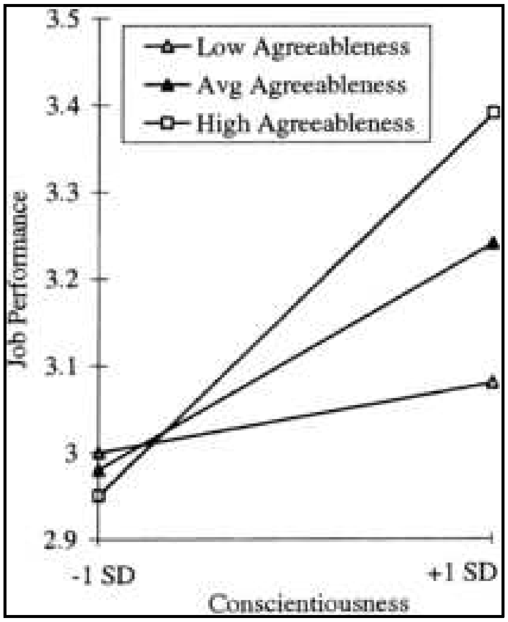
Expertise and Formal Operations Example

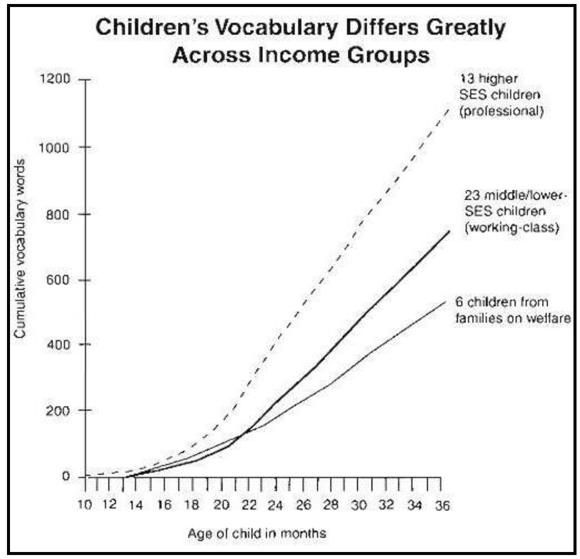
Percent of students displaying formal operational thought

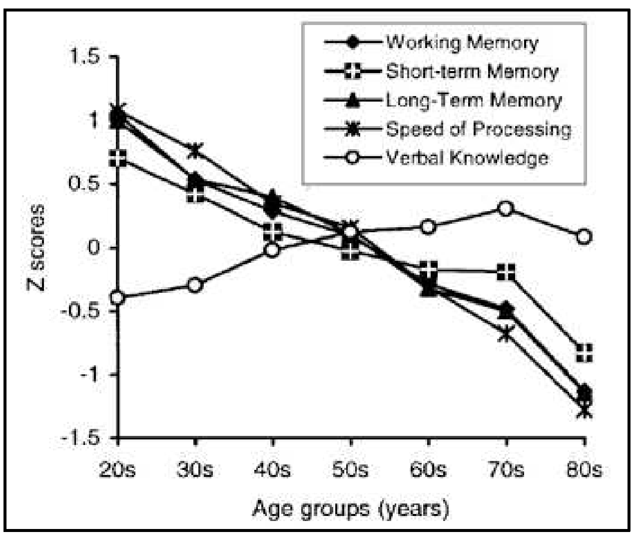
	Physics Majors	Poli Sci Majors	English Majors
Pendulum Problem	90%	50%	40%
Political Problem	60%	80%	40%
Literary Problem	40%	40%	90%

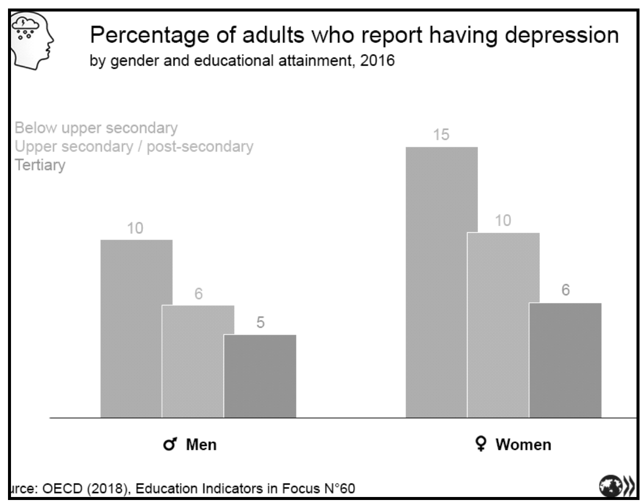
College students show the greatest command of formal operational thought in the subject area most related to their major. (DeLisi & Straudt, 1980)




Low Agreeableness
Avg Agreeableness
High Agreeableness
Job Performance
3.5
3.4
3.3
3.2
3.1
3
2.9
-1 SD
+1 SD
Conscientiousness


Children's Vocabulary Differs Greatly
Across Income Groups
Cumulative vocabulary words
13 higher SES children (professional)
23 middle/lower-SES children (working-class)
6 children from families on welfare
Age of child in months


Working Memory
Short-term Memory
Long-Term Memory
Speed of Processing
Verbal Knowledge
Z scores
20s
30s
40s
50s
60s
70s
80s
Age groups (years)


Percentage of adults who report having depression
by gender and educational attainment, 2016
Below upper secondary
Upper secondary / post-secondary
Tertiary
10
6
5
15
10
6
Men
Women
urce: OECD (2018), Education Indicators in Focus N°60

APPENDIX 4-2: NOTATION FOR FACTORIAL DESIGNS

The general notation for factorial designs is summarized in the following box. The two factors are labelled A and B, respectively; A is the column factor and B the row factor here, but that is arbitrary. Uppercase A and B also refer to the number of levels of each factor (the value represented by k in the single factor designs), and the corresponding lowercase letters act as subscripts (instead of j as in the single factor designs) to represent levels of each factor (i.e., a = 1, 2, ..., A; b = 1, 2, ..., B). We use s (instead of i) and n (or N) to refer to subjects. Capital S denotes the subject variable.

				Factor A						
			1	2	...	a	...	A	Row Means	
		S								
	1	1	y_{111}	y_{211}						
		2	y_{112}							
		s								
		...								
	2	1								
		2								
		s								
		...								
	...									
Factor B	b					y_{abs}				
						$\bar{y}_{ab}$ Cell means			$\bar{y}_b$	
						n_{ab}			n_b	
						SS_{ab}				
						s_{ab}				
	...									
	B									
Column Means						$\bar{y}_a$			$\bar{y}_G$	Grand mean
						n_a			N	
									SS_G	
									s_G	

Notation for Factorial Designs.

Each observation in the design is indicated by y_{abs}, where the three subscripts indicate the level of A, the level of B, and the individual subject (level of S) within that particular AB combination. For example, the fifth observation in the third level of A and the second level of B would be y_{325}. Each cell of the design has n_{ab} observations, representing the number of observations in each combination of a level of A and a level of B. The n_{ab} observations in each cell can be averaged to produce a cell mean, indicated by $\bar{y}_{ab}$. Computing the deviation of each cell observation from the cell mean (i.e., $y-\bar{y}_{ab}$) and summing the squared deviations produces a SSs and standard deviation for each cell, denoted by SS_{ab} and s_{ab}, respectively.

The total number of observations in the analysis is indicated by capital N. Summing all the observations and dividing by N produces a grand mean, represented by $\bar{y}_G$. All the scores could also be used to calculate an SS_G (or SS_{Total}) and standard deviation for all the scores, s_G.

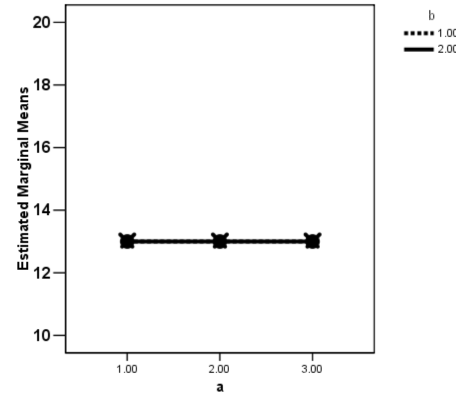
Calculation of the main effect of A requires means and ns for each level of A averaged over levels of B (i.e., $\overline{y}_a$ and n_a). The number of observations n_a will depend on the number of observations in each cell, n_{ab}, and the number of levels of B. The main effect of B will require means and ns for each level averaged across levels of A (i.e., $\overline{y}_b$ and n_b), where n_b will be determined by n_{ab} and the number of levels of A.

APPENDIX 4-3: SAMPLE FACTORIAL ANOVA OUTCOMES

The following analyses illustrate some possible outcomes for a factorial 2 x 3 design. Note independent significance of main effects (a, b) and interaction (a×b). Significant effects are bolded in output. These analyses could also be used to practice calculations using the descriptive statistics.

GLM y1 BY a b /PRINT = DESCR.

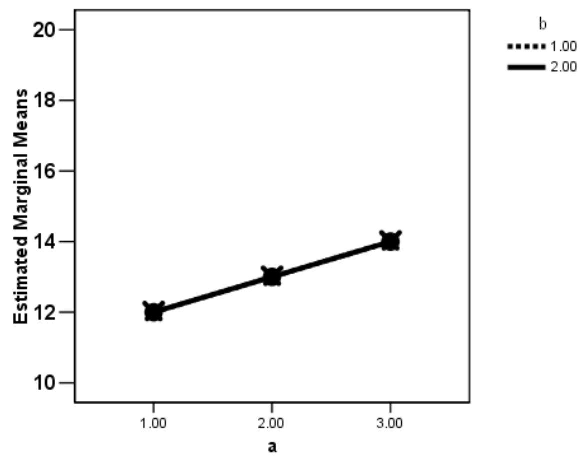
a	b	Mean	Std. Deviation	N
1.00	1.00	13.0000	1.58114	5
	2.00	13.0000	1.58114	5
	Total	13.0000	1.49071	10
2.00	1.00	13.0000	1.58114	5
	2.00	13.0000	1.58114	5
	Total	13.0000	1.49071	10
3.00	1.00	13.0000	1.58114	5
	2.00	13.0000	1.58114	5
	Total	13.0000	1.49071	10
Total	1.00	13.0000	1.46385	15
	2.00	13.0000	1.46385	15
	Total	13.0000	1.43839	30



Source	Type III Sum of Squares	df	Mean Square	F	Sig.
a	.000	2	.000	.000	1.000
b	.000	1	.000	.000	1.000
a * b	.000	2	.000	.000	1.000
Error	60.000	24	2.500		
Corrected Total	60.000	29			

GLM y2 BY a b /PRINT = DESCR.

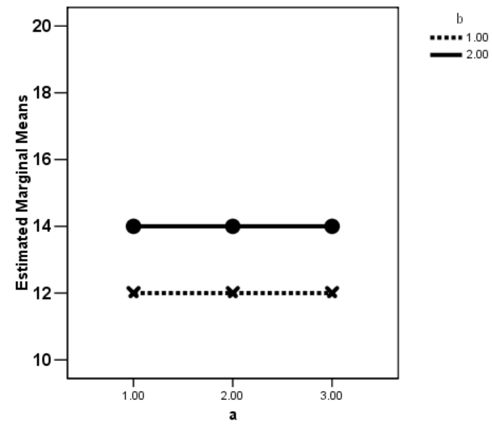
a	b	Mean	Std. Deviation	N
1.00	1.00	12.0000	1.58114	5
	2.00	12.0000	1.58114	5
	Total	12.0000	1.49071	10
2.00	1.00	13.0000	1.58114	5
	2.00	13.0000	1.58114	5
	Total	13.0000	1.49071	10
3.00	1.00	14.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	14.0000	1.49071	10
Total	1.00	13.0000	1.69031	15
	2.00	13.0000	1.69031	15
	Total	13.0000	1.66091	30



Source	Type III Sum of Squares	df	Mean Square	F	Sig.
a	**20.000**	**2**	**10.000**	**4.000**	**.032**
b	.000	1	.000	.000	1.000
a * b	.000	2	.000	.000	1.000
Error	60.000	24	2.500		
Corrected Total	80.000	29			

GLM y3 BY a b /PRINT = DESCR.

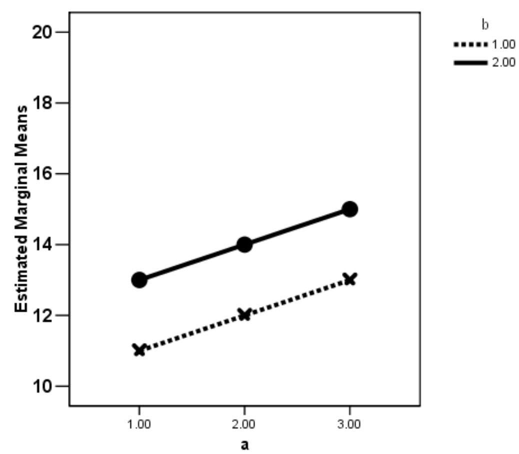
a	b	Mean	Std. Deviation	N
1.00	1.00	12.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	13.0000	1.82574	10
2.00	1.00	12.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	13.0000	1.82574	10
3.00	1.00	12.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	13.0000	1.82574	10
Total	1.00	12.0000	1.46385	15
	2.00	14.0000	1.46385	15
	Total	13.0000	1.76166	30



Source	Type III Sum of Squares	df	Mean Square	F	Sig.
a	.000	2	.000	.000	1.000
b	**30.000**	**1**	**30.000**	**12.000**	**.002**
a * b	.000	2	.000	.000	1.000
Error	60.000	24	2.500		
Corrected Total	90.000	29			

GLM y4 BY a b /PRINT = DESCR.

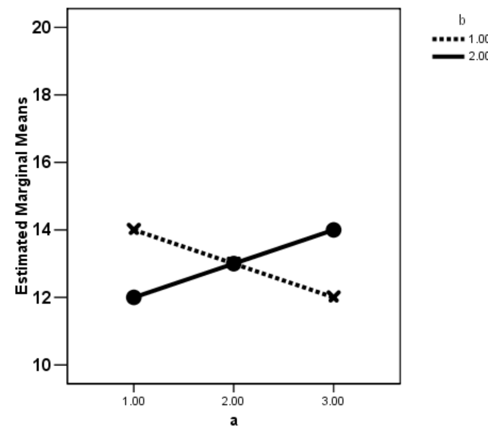
a	b	Mean	Std. Deviation	N
1.00	1.00	11.0000	1.58114	5
	2.00	13.0000	1.58114	5
	Total	12.0000	1.82574	10
2.00	1.00	12.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	13.0000	1.82574	10
3.00	1.00	13.0000	1.58114	5
	2.00	15.0000	1.58114	5
	Total	14.0000	1.82574	10
Total	1.00	12.0000	1.69031	15
	2.00	14.0000	1.69031	15
	Total	13.0000	1.94759	30



Source	Type III Sum of Squares	df	Mean Square	F	Sig.
a	**20.000**	**2**	**10.000**	**4.000**	**.032**
b	**30.000**	**1**	**30.000**	**12.000**	**.002**
a * b	.000	2	.000	.000	1.000
Error	60.000	24	2.500		
Corrected Total	110.000	29			

GLM y5 BY a b /PRINT = DESCR.

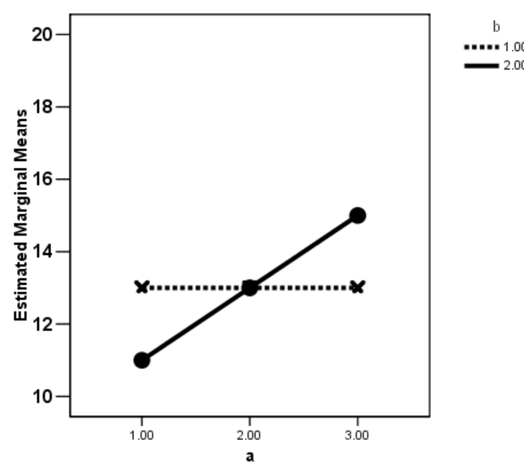
```
a      b     Mean    Std. Deviation N
1.00   1.00  14.0000 1.58114        5
       2.00  12.0000 1.58114        5
       Total 13.0000 1.82574        10

2.00   1.00  13.0000 1.58114        5
       2.00  13.0000 1.58114        5
       Total 13.0000 1.49071        10

3.00   1.00  12.0000 1.58114        5
       2.00  14.0000 1.58114        5
       Total 13.0000 1.82574        10

Total  1.00  13.0000 1.69031        15
       2.00  13.0000 1.69031        15
       Total 13.0000 1.66091        30
```




```
Source          Type III Sum of df Mean Square F        Sig.
                Squares
a               .000            2  .000        .000     1.000
b               .000            1  .000        .000     1.000
a * b           20.000          2  10.000      4.000    .032
Error           60.000          24 2.500
Corrected Total 80.000          29
```

GLM y6 BY a b /PRINT = DESCR.

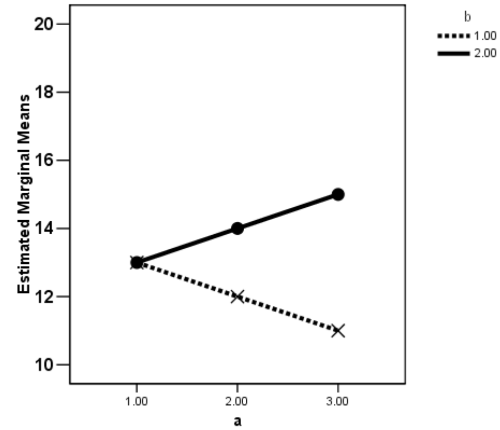
```
a      b     Mean    Std. Deviation N
1.00   1.00  13.0000 1.58114        5
       2.00  11.0000 1.58114        5
       Total 12.0000 1.82574        10

2.00   1.00  13.0000 1.58114        5
       2.00  13.0000 1.58114        5
       Total 13.0000 1.49071        10

3.00   1.00  13.0000 1.58114        5
       2.00  15.0000 1.58114        5
       Total 14.0000 1.82574        10

Total  1.00  13.0000 1.46385        15
       2.00  13.0000 2.23607        15
       Total 13.0000 1.85695        30
```




```
Source          Type III Sum of df Mean Square F        Sig.
                Squares
a               20.000          2  10.000      4.000    .032
b               .000            1  .000        .000     1.000
a * b           20.000          2  10.000      4.000    .032
Error           60.000          24 2.500
Corrected Total 100.000         29
```

GLM y7 BY a b /PRINT = DESCR.

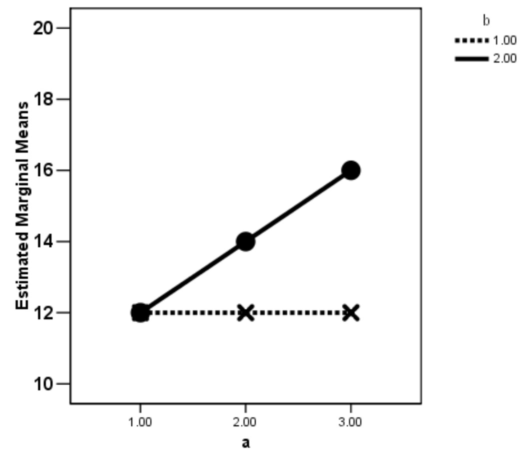
a	b	Mean	Std. Deviation	N
1.00	1.00	13.0000	1.58114	5
	2.00	13.0000	1.58114	5
	Total	13.0000	1.49071	10
2.00	1.00	12.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	13.0000	1.82574	10
3.00	1.00	11.0000	1.58114	5
	2.00	15.0000	1.58114	5
	Total	13.0000	2.58199	10
Total	1.00	12.0000	1.69031	15
	2.00	14.0000	1.69031	15
	Total	13.0000	1.94759	30



Source	Type III Sum of Squares	df	Mean Square	F	Sig.
a	.000	2	.000	.000	1.000
b	**30.000**	**1**	**30.000**	**12.000**	**.002**
a * b	**20.000**	**2**	**10.000**	**4.000**	**.032**
Error	60.000	24	2.500		
Corrected Total	110.000	29			

GLM y8 BY a b /PRINT = DESCR.

a	b	Mean	Std. Deviation	N
1.00	1.00	12.0000	1.58114	5
	2.00	12.0000	1.58114	5
	Total	12.0000	1.49071	10
2.00	1.00	12.0000	1.58114	5
	2.00	14.0000	1.58114	5
	Total	13.0000	1.82574	10
3.00	1.00	12.0000	1.58114	5
	2.00	16.0000	1.58114	5
	Total	14.0000	2.58199	10
Total	1.00	12.0000	1.46385	15
	2.00	14.0000	2.23607	15
	Total	13.0000	2.11725	30



Source	Type III Sum of Squares	df	Mean Square	F	Sig.
a	**20.000**	**2**	**10.000**	**4.000**	**.032**
b	**30.000**	**1**	**30.000**	**12.000**	**.002**
a * b	**20.000**	**2**	**10.000**	**4.000**	**.032**
Error	60.000	24	2.500		
Corrected Total	130.000	29			

APPENDIX 4-4: SAMPLE SIZE & SIGNIFICANCE

The distraction study only had three subjects per condition, which is a very small sample size for any study. Small samples have been used to facilitate calculations and save paper (e.g., for listing data), but sample size is a very important consideration when it comes to significance. With three subjects per cell, the Type by Volume interaction was only marginally significant, p = .052. If we double the sample size to 6 participants per cell, the interaction is now highly significant, p = .00027 (.000 in output). See below.

An important issue in research design is sample size. A very weak (i.e., small) effect may require many subjects for the results to be statistically significant. Or weak effects may benefit from Within-S designs, as discussed in later chapters.

```
DATA LIST FREE / typ vol mis.
BEGIN DATA
1 1 1        1 1 4        1 1 7        1 1 1        1 1 4        1 1 7
1 2 4        1 2 2        1 2 3        1 2 4        1 2 2        1 2 3
1 3 5        1 3 7        1 3 3        1 3 5        1 3 7        1 3 3
1 4 3        1 4 5        1 4 4        1 4 3        1 4 5        1 4 4
2 1 3        2 1 2        2 1 1        2 1 3        2 1 2        2 1 1
2 2 5        2 2 2        2 2 8        2 2 5        2 2 2        2 2 8
2 3 10       2 3 9        2 3 8        2 3 10       2 3 9        2 3 8
2 4 6        2 4 10       2 4 8        2 4 6        2 4 10       2 4 8
END DATA.
GLM mis BY typ vol.
```

Dependent Variable: mis

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
typ	48.000	1	48.000	16.000	.000
vol	120.000	3	40.000	13.333	.000
typ * vol	72.000	3	24.000	8.000	.000
Error	120.000	40	3.000		
Corrected Total	360.000	47			

CHAPTER 5 - FOLLOW-UP ANALYSES FOR FACTORIAL ANOVA

Main effects with $df_{Numerator} > 1$ and interactions, even with df = 1, often require follow-up analyses to determine the specific nature of the differences among means. These follow-up analyses can be post hoc (e.g., all possible pairwise comparisons) or planned. Chapter 5 covers follow-up analyses for the main effects and one approach to the interaction, simple effects analysis. Further follow-up analyses for interactions are covered in Chapter 6.

Post Hoc Comparisons for Main Effects

Below are MANOVA results for the study examining interference by noise or speech (*typ*) as a function of loudness (*vol*). Null hypotheses for the main effects are both rejected. For the main effect of *typ*, df = 1, and therefore the difference must be between the mean for noise, $\overline{y}_{Noise} = 4.0$, and the mean for speech, $\overline{y}_{Speech} = 8.0$; that is, reject H_0: $\mu_{Noise}=\mu_{Speech}$. Speech produced more interference than noise and no further tests are needed. The presence of a marginal interaction, however, should lead us to be cautious about our conclusions for this main effect because the difference between noise and speech may be greater for some levels of the loudness factor. The presence of an interaction complicates interpretation of main effects and generally calls for additional analyses, which are examined later.

Follow-up comparisons are helpful for the *vol* effect given the vague conclusion that one or more of the null hypothesis equalities is false (H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$). There are many ways this H_0 could be false. We need further analysis of variation among the means: $\overline{y}_1 = 3.0$, $\overline{y}_2 = 4.0$, $\overline{y}_3 = 7.0$, $\overline{y}_4 = 6.0$.

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN CELLS	60.00	16	3.75		
typ	24.00	1	24.00	6.40	.022
vol	60.00	3	20.00	5.33	.010
typ BY vol	36.00	3	12.00	3.20	.052
(Model)	120.00	7	17.14	4.57	.006
(Total)	180.00	23	7.83		

The calculation of *t* and *q* for the post hoc tests is similar for the single factor and factorial Between-S designs. The major difference is that n_j for each mean is based on the number of participants at that level of *vol*, summed over levels of the other factor. That is, n_j is the number of observations or scores in the means being compared. In the present case with three scores per cell and two levels of *typ*, $n_j = 3+3 = 6$ for each level of *vol*. Box 5-1 shows the calculation of t for the largest difference

$$t = \frac{7.0-3.0}{\sqrt{3.75(\frac{1}{6}+\frac{1}{6})}} = \frac{4.0}{1.118} = 3.578$$

Box 5-1. Calculation of post hoc t

between means, namely group 1 versus 3. Box 5-2 shows the corresponding calculation for q. For Between-S factorial analyses, MSE from the omnibus ANOVA is used as the standard error.

$$q = \frac{7.0 - 3.0}{\sqrt{3.75(1/6)}} = \frac{4.0}{.791} = 5.057 = 3.578\sqrt{2}$$

Box 5-2. Calculation of post hoc q

For the LSD test, the critical value of t for $df = 16$ is 2.120; therefore, H_0: $\mu_1 = \mu_3$ is rejected. Equivalently, the critical value of q for stretch = 2 is 3.00 (i.e., $t_{Critical} \times \sqrt{2}$), leading to the same conclusion. As for single factor post hoc tests, it helps to lay out the means in a table with rows and columns ordered from the lowest to highest mean. Observed statistics are entered below (these are ts).

Vol		1	2	4	3
	$\bar{y}_V$	3.0	4.0	6.0	7.0
1	3.0	-	0.894	2.683^L	3.578^L
2	4.0		-	1.789	2.683^L
4	6.0			-	0.894
3	7.0				-

```
1      2      4      3
--------      --------
      ---------
```

The summary to the right shows that adjacent groups (12, 24, 43) do not differ significantly; the other three differences (14, 13, 23) are significant. This is a messy conclusion because 1 = 2, 2 = 4, and 4 = 3, but 1 ≠ 4 or 3, and 2 ≠3. The qs are shown below and lead to the same conclusion with stretch = 2.

Vol		1	2	4	3		
	$\bar{y}_V$	3.0	4.0	6.0	7.0	Str	q_α
1	3.0	-	1.264	3.793^{LS}	5.057^{LST}	4	4.05
2	4.0		-	2.528	3.793^{LS}	3	3.65
4	6.0			-	1.264	2	3.00
3	7.0				-		

```
Summaries          1      2      4      3
      LSD & SNK    ----------   ----------
                          ----------

      Tukey        -------------------
                          -------------------
```

For the Tukey procedure, the stretch of 4 is used for all comparisons. Only the largest difference is significant, that between groups 3 and 1. As shown in the summary, one subset of groups that do not differ significantly contains groups 1, 2, and 4, and a second subset of groups that do not differ significantly contains groups 2, 4, and 3. Five of the comparisons are not significant: 12, 14, 24 from subset one and 24 (again), 23, and 43 from subset two. Only the 1 vs 3 comparison is significant. This is an awkward conclusion because 1 = 2 = 4 and 2 = 4 = 3, but 1 ≠ 3.

The significance of qs for the SNK procedure requires three critical values, one benefit of arranging the comparisons as above. A stretch of 2 and critical q = 3.00 is used for the three adjacent comparisons that span only two groups when means are ordered: 12, 24, 43. Observed qs for these comparisons are on the diagonal: 1.264, 2.525, and 1.264. They are not significant. A stretch of 3 and critical q = 3.65 is used for 14 and 23, which span three groups. The observed qs for these comparisons appear immediately up from the diagonal: 3.793 and 3.793. Both are significant. Finally, q = 5.057 for the comparison spanning all four groups, 13, is in the upper right corner, and is significant.

In practice, SNK tests are done starting with the upper right cell, working left and down, and stopping when a nonsignificant comparison occurs. That is, first test 5.057, which is significant, so continue to 3.793 to its left. This comparison is significant, so continue to the next cell to the left, which is not significant. Notice if the 3.793 was not significant and we had tested the next cell, it is possible that its value would exceed 3.00, the lower critical value of q for that cell. For example, the q for the 12 comparison could have been 3.20. Stopping at the first nonsignificant result on a row avoids the paradox of a smaller difference being significant when a larger one is not.

After stopping in the first row, start at the right of the second row because the comparison above is significant, and continue in the same manner. The right-most entry in row two (q_{23} = 3.793) is significant, but the next entry to the left is not. Because the 23 comparison was significant, test the cell below that; if the 23 comparison was not significant, we would not test the cell below because it could be significant despite involving a smaller difference between means. In the present study, the SNK procedure leads to the same conclusions as the LSD procedure.

The Bonferroni procedure is difficult without computer assistance. We would need critical values for t (or q) for α = .05/6 = .0083, which is not generally found in tables. SPSS could be used to generate these critical values. Alternatively, we can obtain the exact p values for the t statistic (or q with stretch = 2) and compare them to .0083, or equivalently, compare $6 \times p_{LSD}$ to .05.

The GLM below performs the four tests and the results correspond to the preceding calculations: SE = 1.118 (the denominator for the t test), the same pairwise comparisons are significant, and the Homogeneous Subsets of groups that do not differ significantly are identical to our underlined summaries. Also the Bonferroni p values correspond to the LSD p values multiplied by the number of comparisons, 6, and the tests become more conservative from LSD, to SNK, to TUKEY, to Bonferroni.

```
GLM  mis  BY typ vol  /POSTHOC = vol ( LSD SNK TUKEY BONFERRONI ).
...
            (I)    (J)    Mean Difference Std.      Sig.
            vol    vol    (I-J)           Error
 LSD        1.0000 2.0000 -1.000000       1.1180340 .384
                   3.0000 -4.000000(*)    1.1180340 .003
                   4.0000 -3.000000(*)    1.1180340 .016
            2.0000 3.0000 -3.000000(*)    1.1180340 .016
                   4.0000 -2.000000       1.1180340 .093
            3.0000 4.0000 1.000000        1.1180340 .384

 Tukey HSD  1.0000 2.0000 -1.000000       1.1180340 .808
                   3.0000 -4.000000(*)    1.1180340 .012
                   4.0000 -3.000000       1.1180340 .070
            2.0000 3.0000 -3.000000       1.1180340 .070
                   4.0000 -2.000000       1.1180340 .314
            3.0000 4.0000 1.000000        1.1180340 .808

 Bonferroni 1.0000 2.0000 -1.000000       1.1180340 1.000
                   3.0000 -4.000000(*)    1.1180340 .015
                   4.0000 -3.000000       1.1180340 .098
            2.0000 3.0000 -3.000000       1.1180340 .098
                   4.0000 -2.000000       1.1180340 .556
            3.0000 4.0000 1.000000        1.1180340 1.000

Homogeneous Subsets
                           1        2        3
 Student-Newman- 1.0000 6 3.000000
 Keuls(a,b,c)    2.0000 6 4.000000 4.000000
                 4.0000 6          6.000000 6.000000
                 3.0000 6                   7.000000
                 Sig.     .384     .093     .384

 Tukey           1.0000 6 3.000000
 HSD(a,b,c)      2.0000 6 4.000000 4.000000
                 4.0000 6 6.000000 6.000000
                 3.0000 6          7.000000
                 Sig.     .070     .070
```

Although all four post hoc tests are shown for learning purposes, in practice researchers decide how conservative to be prior to seeing the results and use only the post hoc procedure that provides the desired control of a Type I error across all comparisons. Post hoc tests also require that the omnibus F be significant for the factor being examined. Here, *vol* was significant.

The post hoc procedures did not lead to a tidy conclusion about the relationship between *vol* and mistakes, nor do pairwise comparisons allow for more sophisticated hypotheses to be tested. It might reasonably be expected, for example, that mistakes would increase as volume increased. This linear contrast can be tested using a priori or planned comparisons.

A Priori or Planned Comparisons for Main Effects

Planned comparisons are calculated much as for the single factor design (see Box 5-3), but again using the appropriate n in the

$$L = \sum c_j \times \bar{y}_j \qquad SS_L = \frac{n_j \times L^2}{\sum c_j^2} \qquad F = \frac{\frac{SS_L}{1}}{MS_{Error}}$$

Box 5-3. Contrast Formula

calculations. The proper n will depend on the number of observations per cell, and the number of cells across which the main effects are averaged. As for single factor designs, planned comparisons can be carried out even if the omnibus F for the main effect is not significant. Polynomial planned comparisons for *vol* are calculated below given volume is an ordered factor. MS_{Error} = 3.75 from the omnibus ANOVA.

	Vol	1	2	3	4				
	$\bar{y}_V$	3.0	4.0	7.0	6.0	L	SS	$= n_v \times L^2/\Sigma c_j^2$	F
Linear		−3	−1	+1	+3	12.0	43.20	$= 6\times 12^2/20$	11.52
Quadratic		+1	−1	−1	+1	−2.0	6.00		1.60
Cubic		−1	+3	−3	+1	−6.0	10.80		2.88
						Σ =	60.00	$= SS_{Vol}$	$\bar{F} = 5.333 = F_{Vol}$

With df = 1, 16 and alpha = .05, $F_{Critical}$ = 4.49. We reject the null hypothesis of no linear relationship between mistakes and volume, and conclude there is a linear relationship. Specifically, mistakes increase significantly as volume increases. If a directional test were appropriate because an increase in mistakes was predicted and not a decrease, then we could justify using $F_{Critical}$ = 3.05 (i.e., F for a = .10) and dividing the observed p value by two. The quadratic and cubic effects are not significant.

Given k - 1 orthogonal contrasts, SSs for the contrasts sum to SS for the main effect and the mean of the Fs equals the omnibus F. These relationships illustrate why planned contrasts can be significant even when the omnibus F is not. Instead of dividing SS = 60.0 equally across df = 3, planned contrasts allow much of the variability to load on a single df, producing a larger $MS_{Numerator}$ and larger F. Moreover, the critical value of F for df = 1, 16 is less than the critical value for df = 3, 16.

Contrasts can also be tested for significance using the t-test in Box 5-4. The conclusions are identical as the tests are equivalent.

$$t_{Linear} = \frac{L-0}{\sqrt{MSE\times\Sigma(\frac{c_j^2}{nj})}} = \frac{12.0-0}{\sqrt{3.75\times(\frac{-3^2}{6}+\frac{-1^2}{6}+\frac{1^2}{6}+\frac{3^2}{6})}} = \frac{12.0}{3.536} = 3.394 = \sqrt{11.52}$$

Box 5-4. Calculation of t-test for Linear contrast

Recall that SPSS sometimes normalizes contrasts or performs some other transformation. Here are calculations for the linear contrast using normalized coefficients; that is, integer coefficients divided by $\sqrt{\Sigma c_j^2} = \sqrt{20}$. The sum of normalized coefficients squared is 1.0, which is the denominator for the calculation of SS_{Lin}. The corresponding t-test using normalized coefficients equals that in Box 5-4.

					L	SS
Linear	−.6708	−.2236	.2236	.6708	2.6832	$43.20 = 6\times 2.6832^2$
						$\approx 12.0/\sqrt{20}$

SPSS and Planned Contrasts for Main Effects

As with the single factor design, SPSS provides diverse ways to perform contrasts. The first analysis uses MANOVA and the SINGLEDF option. In the summary table, the main effect of *vol* is partitioned into linear (1st parameter), quadratic, and cubic components because POLYNOMIAL is stated as the contrast. Only the linear contrast is significant. T-tests that correspond to the single df F tests appear following the summary table. Note that the values given for L and SE are normalized coefficients, rather than the integer values we used initially. The conclusions are the same. SINGLEDF also partitions the type by volume interaction, which is examined later.

```
MANOVA mis BY typ(1 2) vol(1 4) /CONTRAST(vol) = POLY /PRINT = SIGNI(SINGLEDF).
```

```
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              60.00      16      3.75
 typ                       24.00       1     24.00      6.40      .022
 vol                       60.00       3     20.00      5.33      .010
   1ST Parameter           43.20       1     43.20     11.52      .004
   2ND Parameter            6.00       1      6.00      1.60      .224
   3RD Parameter           10.80       1     10.80      2.88      .109
 typ BY vol                36.00       3     12.00      3.20      .052
   1ST Parameter           30.00       1     30.00      8.00      .012
   2ND Parameter            6.00       1      6.00      1.60      .224
   3RD Parameter             .00       1       .00       .00     1.000

 (Model)                  120.00       7     17.14      4.57      .006
 (Total)                  180.00      23      7.83

typ
  Parameter          Coeff.    Std. Err.          t-Value           Sig. t
        2     -1.0000000000       .39528         -2.52982           .02229

 vol
  Parameter          Coeff.    Std. Err.          t-Value           Sig. t
        3      2.6832815730       .79057          3.39411           .00371
        4     -1.0000000000       .79057         -1.26491           .22402
        5     -1.3416407865       .79057         -1.69706           .10905

 typ BY vol
  Parameter          Coeff.    Std. Err.          t-Value           Sig. t
        6     -2.2360679775       .79057         -2.82843           .01211
        7      1.0000000000       .79057          1.26491           .22402
        8       .0000000000       .79057           .00000          1.00000
```

The next analysis shows GLM results for POLYNOMIAL contrasts. The results agree with earlier analyses and again show normalized coefficients. This is important to remember if SSs are calculated by hand, as shown below. Note that n_jL^2 is not divided by $\sum c^2_j$ (technically it is divided by 1.0). The ANOVA that follows with df = 3 is redundant with the main effect.

```
GLM mis BY typ vol /CONTRAST(vol) = POLY.
...
Custom Hypothesis Tests

 Linear         Contrast Estimate           2.683    SS = 6 × 2.683² = 43.19
                Std. Error                  .791      t = 2.683/.791 = 3.392
                Sig.                        .004
 Quadratic      Contrast Estimate           -1.000
                Std. Error                  .791
                Sig.                        .224
 Cubic          Contrast Estimate           -1.342
                Std. Error                  .791
                Sig.                        .109

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 60.000         3  20.000      5.333 .010
 Error    60.000         16 3.750
```

Modifying GLM to specify separate integer contrasts produces the following results. Now the contrasts correspond to our calculations, and the ANOVAs provide $SS_{Contrast}$ for each contrast. These SSs show the partitioning of SS_{Vol} and allow for calculating η^2 for each contrast.

```
GLM mis BY typ vol /CONTRAST(vol) = SPECIAL(-3 -1  1  3)
                   /CONTRAST(vol) = SPECIAL( 1 -1 -1  1)
                   /CONTRAST(vol) = SPECIAL(-1  3 -3  1).
...
Custom Hypothesis Tests #1
 L1          Contrast Estimate           12.000      SS = 6×12.0²/20 = 43.2
             Std. Error                  3.536       t = 12.0/3.536 = 3.394
             Sig.                        .004

 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 43.200         1  43.200      11.520 .004
 Error    60.000         16 3.750

Custom Hypothesis Tests #2
 L1          Contrast Estimate           -2.000
             Std. Error                  1.581
             Sig.                        .224

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 6.000          1  6.000       1.600 .224
 Error    60.000         16 3.750

Custom Hypothesis Tests #3
 L1          Contrast Estimate           -6.000
             Std. Error                  3.536
             Sig.                        .109

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 10.800         1  10.800      2.880 .109
 Error    60.000         16 3.750
```

Although polynomial contrasts make sense for Volume, other planned contrasts are possible. Prior research or theory might indicate that sounds below a certain threshold are minimally distracting, whereas sounds above the threshold interfere and lead to more mistakes. For a threshold between volumes 2 and 3, appropriate contrasts might be as follows (t-test results have been deleted). The first contrast (1&2 vs 3&4) captures more variability than the linear contrast (54.0 vs 43.2).

```
MANOVA mis BY typ(1 2) vol(1 4)
 /CONTRAST(vol) = SPECIAL(1 1 1 1  -1 -1 1 1  -1 1 0 0  0 0 -1 1)
 /DESIGN typ vol(1) vol(2) vol(3) typ BY vol.
```

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	60.00	16	3.75		
TYP	24.00	1	24.00	6.40	.022
VOL(1)	**54.00**	**1**	**54.00**	**14.40**	**.002**
VOL(2)	3.00	1	3.00	.80	.384
VOL(3)	3.00	1	3.00	.80	.384
TYP BY VOL	36.00	3	12.00	3.20	.052
(Model)	120.00	7	17.14	4.57	.006
(Total)	180.00	23	7.83		

Three final observations about main effects. First, it is possible to carry out main effect contrasts by using the cell means, rather than row or column means. The contrast coefficients are repeated for each level of the *other* factor and n_j is adjusted accordingly. Here are calculations for the linear contrast.

	Noise				Speech				
	1	2	3	4	1	2	3	4	
$\bar{y}_{ab}$	4.0	3.0	5.0	4.0	2.0	5.0	9.0	8.0	L
Lin	−3	−1	1	3	−3	−1	1	3	24.0

$SS_{Linear} = 3 \times 24^2/40 = 43.20$

In essence the main effect contrast tests whether the linear pattern relating mistakes to volume is the same for both levels of type and is significant averaged over type. The "averaging" was done earlier by using column means averaged over the Noise and Speech conditions, and is done here by calculating the contrast coefficient L with the same linear coefficients repeated for the Noise and Speech conditions.

The second point is that some contrasts, but not all, can be conceptualized as differences between means or variability among means averaged over groups. To illustrate, consider the -1 -1 +1 +1 contrast shown in the preceding MANOVA. Contrast calculations can be done as follows.

Vol	1	2	3	4		
$\bar{y}_V$	3.0	4.0	7.0	6.0	L	$SS = n_j \times L^2 / \Sigma c_j^2$
C12v34	−1	−1	+1	+1	6.0	$54.0 = 6 \times 6^2 /4$

Alternatively, we could group means coded the same on contrasts and calculate SS for the difference between the means, as shown below.

	1&2	3&4	
$\bar{y}_V$	3.50	6.50	$\bar{y}_G = 5.0$
n_j	12	12	$N = 24$
$\bar{y}_V - \bar{y}_G$	−1.50	+1.50	

```
SS = 12(-1.50² + 1.50²) = 54.0
```

Here is a second example where contrast coefficients correspond to a difference between means. When would this contrast be appropriate? What orthogonal contrasts are possible?

```
Vol          1     2     3     4
ȳV           3.0   4.0   7.0   6.0          L     SS    = nj x L² / Σcj²
C1v234       -3    +1    +1    +1           8.0   32.0  = 6 x 8² /12

             1           2&3&4
ȳV           3.0         5.667       ȳG = 5.0
nj           6           18          N = 24
ȳV - ȳG      -2.0        +.667

SS = 6 × -2.0² + 18 × .667² = 32.0
```

One final observation. GLM can test contrasts using the /LMATRIX option but the main effect contrast must also be coded along with the interaction, similar to what was just done with contrasts using cell means. The initial part of the command is straightforward; /LMATRIX is followed by the name of the factor and then the contrast coefficients. Ideally this would work, but it does not. SPSS would report that the L Matrix is not estimable and produce blank results.

GLM requires that the contrast be repeated for the interaction term as well. So *factor BY factor* is followed by the coefficients again, but now divided by the number of levels of the *other* factor (*typ* has two levels here). As examined more later, it is also critical that the initial GLM command (i.e., GLM *mis* BY *typ vol*) specify the factors in the correct order or the analysis will either not run or provide incorrect results. That is, BY *typ vol* differs from BY *vol typ*. If one gets over all these hurdles, the results will agree with those obtained earlier, as shown below.

```
GLM mis BY typ vol /LMATRIX vol        -3   -1    1    3
                            vol BY typ -3/2 -1/2  1/2  3/2
                                       -3/2 -1/2  1/2  3/2.
...
Custom Hypothesis Tests
 L1       Contrast Estimate          12.000
          Std. Error                 3.536
          Sig.                       .004

 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 43.200         1  43.200      11.520 .004
 Error    60.000         16 3.750
```

This concludes follow-up analyses for the main effects of a factorial design. We next examine one approach to follow-up analyses for interactions, simple effects.

Follow-Up Analyses for Interactions: Simple Effects

When an interaction is predicted, follow-up analyses will determine whether the interaction corresponds to the expected pattern. Even when no interaction is predicted, follow-up analyses may be called for when the interaction effect is significant, or sometimes when it is marginally significant as in the present study with $p = .052$ for the *typ* by *vol* interaction. Marginally significant interactions warrant closer examination because the omnibus F for interactions is not very sensitive except for cross-over interactions where there are no main effects. That is, one factor has opposite effects across levels of the other factor. This is seldom the only pattern of interest. When the interaction is not pure cross-over, some of the variability due to the interaction will be allocated to one or both main effects.

There are two approaches to follow-up analyses of the interaction. One approach examines the effects of one factor within levels of the other factor; for example, the effect of Type at each of the four levels of Volume, or the effect of Volume at each level of Type. This is called a simple effects analysis. A second approach described in Chapter 6 partitions the interaction into planned components.

One way to conceptualize the simple effects analysis is as a single factor ANOVA repeated at each level of the other factor. That is, each level of one factor is treated as a separate study, except that MS_{Error} from the omnibus ANOVA is the denominator when all factors are Between-S. To calculate SS for simple effects, cell means are subtracted from row or column means, depending on which simple effect analysis is done. Row and column means function as the grand mean. Here are calculations for the simple effect of volume within levels of type.

	Volume (B)				
Type (A)	**1.**	**2.**	**3.**	**4.**	$\bar{y}_a$
1. Noise					
$\bar{y}_{ab}$	4.0	3.0	5.0	4.0	$\bar{y}_{1.}$ 4.0
$\bar{y}_{ab}-\bar{y}_{1.}$	0.0	−1.0	+1.0	0.0	$SS_{VwN} = 3(0^2+-1^2+1^2+0^2) = 6.0$
					df = 4 − 1 = 3
2. Speech					
$\bar{y}_{ab}$	2.0	5.0	9.0	8.0	$\bar{y}_{2.}$ 6.0
$\bar{y}_{ab}-\bar{y}_{2.}$	−4.0	−1.0	+3.0	+2.0	$SS_{VwS} = 3(-4^2+-1^2+3^2+2^2) = 90.0$
					df = 4 − 1 = 3

Clearly the effect of volume on mistakes is much stronger for Speech as shown by earlier graphs. There is little variability among the four means for the Noise stimuli. Indeed, the F for Noise is not significant, whereas that for Speech is ($F_{Critical} = 3.24$). The “w” in “VwN” stands for “Within.” Simple effects tests the effect of one factor within the levels of the other factor.

$F_{VwN} = (6.0/3)/3.75 = 2.0/3.75 = .5333$ Do not reject H_0: $\mu_{11} = \mu_{12} = \mu_{13} = \mu_{14}$

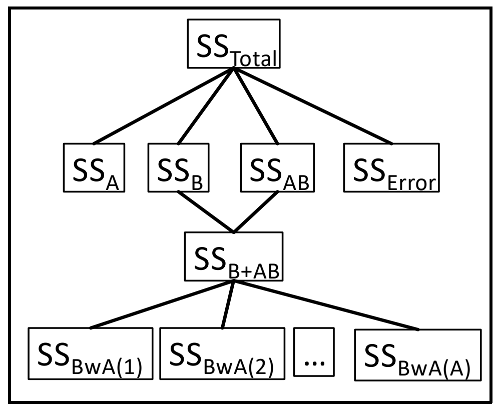
$F_{VwS} = (90.0/3)/3.75 = 30.0/3.75 = 8.00$ Reject H_0: $\mu_{21} = \mu_{22} = \mu_{23} = \mu_{24}$

Both MANOVA and GLM include methods to carry out simple effects analyses. MANOVA uses the /DESIGN option with the term WITHIN (or just W) to define the simple effect of one factor at each level of the other factor. The following commands request that SS_{Total} be partitioned into SS_{Typ}, $SS_{VwT(1)}$, and $SS_{VwT(2)}$. On the design option, *typ(1)* following WITHIN refers to level one of the *typ* factor (i.e., the Noise condition), and *typ(2)* refers to level two of *typ* (i.e., the Speech condition). SSs agree with our calculations and the effect of *vol* is significant only for the Speech condition.

```
MANOVA mis BY typ(1 2) vol(1 4)   /DESIGN typ vol WITHIN typ(1) vol WITHIN typ(2).
```

```
 Source of Variation          SS       DF        MS         F  Sig of F
 WITHIN+RESIDUAL           60.00       16      3.75
 TYP                       24.00        1     24.00      6.40      .022
 VOL WITHIN TYP(1)          6.00        3      2.00       .53      .666
 VOL WITHIN TYP(2)         90.00        3     30.00      8.00      .002

 (Model)                  120.00        7     17.14      4.57      .006
 (Total)                  180.00       23      7.83
```

The simple effects analysis does *not* partition the interaction. Rather, simple effects analyses are an alternative to the default partitioning into SS_T, SS_V, and SS_{TV}. Specifically, simple effects analyses partition variability among the means into the main effect for one factor and the simple effects for the other factor. The bottom cells in Box 5-5 show the alternative partitioning. The present analysis, for example, involves the main effect of *typ* and simple effects of *vol* at each of two levels of *typ*. The simple effects analysis for *typ* within *vol* involves a main effect for *vol* and simple effects for *typ* at each level of *vol*. In general,



Box 5-5. Alternative Partitioning for Default and Simple Effects Analyses.

$SS_{Total} = SS_A + SS_B + SS_{AxB} + SS_E = SS_A + SS_{BwA(1)} + SS_{BwA(2)} ... + SS_{BwA(A)} + SS_E$

Therefore, $SS_{BwA(1)} ... + SS_{BwA(A)} = SS_B + SS_{AxB}$

In the present study,

$SS_{VwT(1)} + SS_{VwT(2)} = 6.0 + 90.0 = 96.0 = SS_V + SS_{TxV} = 60.0 + 36.0$

$df_{VwT(1)} + df_{VwT(2)} = 3 + 3 = 6 = df_V + df_{TxV} = 3 + 3$

Some statisticians object to simple effects analyses because main effects and interaction are confounded, as shown above and in Box 5-5. But in many cases, simple effects provide a more sensible approach to the interaction than the standard interaction effect. Here, for example, we have markedly different effects of volume for Noise and Speech despite the interaction effect being only marginally significant ($p = .052$). The omnibus F for the interaction is marginal because variability due mostly to the simple effect for Speech is attributed to main effects and not just to the interaction.

Simple effects analyses in SPSS must be done carefully. In particular, only the full partitioning shown in Box 5-5 will ensure that the correct error term is used. The MANOVA shown above includes the main effect of type as well as the two simple effects. This guarantees that $MS_{\text{Error}} = 3.75 = MS_{\text{Error}}$ from the factorial ANOVA. The following (incorrect) analysis shows that when the main effect of type is omitted, $MS_{\text{Error}} = 4.94$, which includes SS_{Type} and its df = 1. So the /DESIGN statement must be specified correctly or other approaches used to ensure the correct error term.

```
MANOVA mis BY typ(1 2) vol(1 4)   /DESIGN vol WITHIN typ(1) vol WITHIN typ(2).
 Source of Variation            SS       DF         MS          F  Sig of F
 WITHIN+RESIDUAL             84.00       17       4.94 Wrong Error Term!!
 VOL WITHIN TYP(1)            6.00        3       2.00        .40      .751
 VOL WITHIN TYP(2)           90.00        3      30.00       6.07      .005
```

GLM uses the /EMMEANS or the /LMATRIX options to conduct simple effects analyses. The Estimated Marginal Means option calculates means for main effects or factorial cells using the TABLE(factor) or TABLE(factor BY factor) subcommand. Following the factorial version, COMPARE(factor) produces the simple effects result. In addition, COMPARE performs pairwise comparisons between levels of the simple effects factor within levels of the other factor, as seen below.

```
GLM mis BY typ vol /EMMEANS TABLE(typ BY vol) COMPARE(vol).
...
Estimated Marginal Means
typ * vol
Pairwise Comparisons
 typ    (I)    (J)    Mean Difference Std.  Sig.(a)
        vol    vol    (I-J)           Error

 1.0000 1.0000 2.0000 1.000           1.581 .536
               3.0000 -1.000          1.581 .536
               4.0000 .000            1.581 1.000     No pairwise comparisons
        2.0000 3.0000 -2.000          1.581 .224      significant for Noise
               4.0000 -1.000          1.581 .536
        3.0000 4.0000 1.000           1.581 .536

 2.0000 1.0000 2.0000 -3.000          1.581 .076      NS
               3.0000 -7.000(*)       1.581 .000      Sig.
               4.0000 -6.000(*)       1.581 .002      Sig.
        2.0000 3.0000 -4.000(*)       1.581 .022      Sig.
               4.0000 -3.000          1.581 .076      NS
        3.0000 4.0000 1.000           1.581 .536      NS
```

typ		Sum of Squares	df	Mean Square	F	Sig.
1.0000	Contrast	6.000	3	2.000	.533	.666
	Error	60.000	16	3.750		
2.0000	**Contrast**	**90.000**	**3**	**30.000**	**8.000**	**.002**
	Error	60.000	16	3.750		

The final summary tables show simple effects results identical to those calculated and produced by MANOVA. The pairwise comparisons shown above the summary tables do not adjust for the number of comparisons within a level of *typ*, and correspond to LSD tests. For the Noise condition, none of the differences are significant, not surprising since the omnibus *F* is far from significant. For the Speech condition, level 1 differs significantly from 3 and 4, 2 differs from 3, 1 and 2 are marginally different, as are 2 and 4; 3 and 4 do not differ significantly. Reported as subsets of conditions not differing significantly, the results could be summarized as:

```
                 1       2       4       3
Noise            ---------------------------
Speech           ----------      ----------
                         ----------
```

EMMEANS can produce Bonferroni p values by following COMPARE with ADJ(BONFERRONI); the values would be 6 times the *p* values reported above. Other post hoc procedures are not available, although the statistics above could be used to calculate *t*s, which could be multiplied by √2 to produce *q*s that could be compared to appropriate critical values for the SNK and Tukey tests. For example, $t_{12} = 3.00/1.581 = 1.898$, and $q_{12} = 1.898 \times \sqrt{2} = 2.684$.

Planned comparisons for simple effects are considered in Chapter 6, along with partitioning the interaction, a second approach to follow-up analyses for interactions.

CHAPTER 6 - MORE ON FOLLOW-UP ANALYSIS OF INTERACTIONS

The final topics for follow-up analyses of Between-S factorials are planned comparisons for simple effects and partitioning the interaction.

Planned Comparisons for Simple Effects

Rather than post hoc comparisons, simple effects analyses can be followed by planned comparisons. Although these can be calculated ignoring one level of the other factor, there is a conceptual benefit to including all cells and using 0s for some coefficients, as shown below. Recall that calculation of the linear contrast for the main effect using cell means repeated the linear coefficients for the Noise and Speech conditions. For simple effects contrasts, 0s are used for one level of the second factor. Calculations are shown for the linear effect of volume within the Speech condition using our standard formula for contrasts and the appropriate n_j for each cell mean; here $n_j = 3$.

		Noise				Speech					
		1	2	3	4	1	2	3	4		
	$\bar{y}_{ab}$	4.0	3.0	5.0	4.0	2.0	5.0	9.0	8.0	L	SS
Vol Within Noise											
	Lin	−3	−1	1	3	0	0	0	0		
	Qua	1	−1	−1	1	0	0	0	0		
	Cub	−1	3	−3	1	0	0	0	0		
Vol Within Speech											
	Lin	0	0	0	0	−3	−1	1	3	22.0	$72.60 = 3\times22^2/20$
	Qua	0	0	0	0	1	−1	−1	1		
	Cub	0	0	0	0	−1	3	−3	1		

Planned comparisons for simple effects are easier in MANOVA than GLM. In MANOVA, specifying the contrasts produces default t-test results, and with other SPSS commands also partitions the SSs for the simple effects and reports Fs.

In the following analysis, the 1st, 2nd, and 3rd parameters in the summary table refer to the Linear, Quadratic, and Cubic contrasts within levels of the *typ* factor. SS for the Linear effect of *vol* within *typ(2)* (i.e., Speech) agrees with our calculations and is highly significant, $F = 19.36$, $p = .000$. None of the effects for *vol* within *typ(1)* (i.e., Noise) are close to significant. The following t-tests are redundant with the *F* tests, but show again that MANOVA uses normalized contrast coefficients since the Parameter 6 coefficient of 4.9193 equals our L of 22.0 divided by $\sqrt{20}$, which is $\sqrt{\Sigma c_j^2}$.

```
MANOVA mis BY typ(1 2) vol(1 4) /CONTRAST(vol) = POLY /PRINT = SIGNI(SINGLEDF)
  /DESIGN typ vol WITHIN typ(1) vol WITHIN typ(2).

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           60.00      16      3.75
 TYP                       24.00       1     24.00      6.40      .022
 VOL WITHIN TYP(1)          6.00       3      2.00       .53      .666
   1ST Parameter             .60       1       .60       .16      .694
   2ND Parameter             .00       1       .00       .00     1.000
   3RD Parameter            5.40       1      5.40      1.44      .248
 VOL WITHIN TYP(2)         90.00       3     30.00      8.00      .002
   1ST Parameter           72.60       1     72.60     19.36      .000
   2ND Parameter           12.00       1     12.00      3.20      .093
   3RD Parameter            5.40       1      5.40      1.44      .248
...
 VOL WITHIN TYP(2)
  Parameter           Coeff.     Std. Err.   t-Value     Sig. t
        6       4.9193495505       1.11803   4.40000     .00045
        7      -2.0000000000       1.11803  -1.78885     .09259
        8      -1.3416407865       1.11803  -1.20000     .24761
```

Contrast effects can also be specified on the /DESIGN option. The following analyis performs the polynomial analysis using the /DESIGN statement, but now with integer coefficients rather than POLYNOMIAL. Numbers in brackets after *vol* refer to linear, quadratic, and cubic contrasts. Numbers after *typ* refer to levels of the factor. For example, "vol(1) W typ(2)" denotes the linear effect of *vol* for the second level of *typ* (i.e., Speech). The integer contrast coefficients produce an L that corresponds to our earlier calculations, whereas POLYNOMIAL led to a normalized value. T-tests for the Noise contrasts are deleted.

```
MANOVA mis BY typ(1 2) vol(1 4)
 /CONTRAST(vol) = SPECIAL(1 1 1 1  -3 -1 1 3  1 -1 -1 1  -1 3 -3 1)
 /DESIGN typ vol(1) W typ(1) vol(2) W typ(1)  vol(3) W typ(1)
             vol(1) W typ(2) vol(2) W typ(2)  vol(3) W typ(2).
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           60.00      16      3.75
 TYP                       24.00       1     24.00      6.40      .022
 VOL(1) W TYP(1)             .60       1       .60       .16      .694
 VOL(2) W TYP(1)             .00       1       .00       .00     1.000
 VOL(3) W TYP(1)            5.40       1      5.40      1.44      .248
 VOL(1) W TYP(2)           72.60       1     72.60     19.36      .000
 VOL(2) W TYP(2)           12.00       1     12.00      3.20      .093
 VOL(3) W TYP(2)            5.40       1      5.40      1.44      .248
...
 VOL(1) W TYP(2)
  Parameter           Coeff.        Std. Err.          t-Value           Sig. t
        6      22.0000000000          5.00000          4.40000           .00045
 VOL(2) W TYP(2)
  Parameter           Coeff.        Std. Err.          t-Value           Sig. t
        7      -4.0000000000          2.23607         -1.78885           .09259
 VOL(3) W TYP(2)
  Parameter           Coeff.        Std. Err.          t-Value           Sig. t
        8      -6.0000000000          5.00000         -1.20000           .24761
```

The following analysis shows results for a set of contrasts consistent with a threshold model; that is, sounds louder than level 2 will affect performance. Numbers in parentheses after *vol* and before W again refer to contrasts, but now vol(1) refers to contrast -1 -1 1 1. Numbers after W (WITHIN) and *typ* again refer to the levels of that factor; that is, typ(1) refers to the Noise condition.

```
MANOVA mis BY typ(1 2) vol(1 4)
 /CONTRAST(vol) = SPECIAL(1 1 1 1  -1 -1 1 1  -1 1 0 0  0 0 -1 1)
 /DESIGN typ vol(1) W typ(1) vol(2) W typ(1)  vol(3) W typ(1)
             vol(1) W typ(2) vol(2) W typ(2)  vol(3) W typ(2).

 Source of Variation          SS       DF        MS          F  Sig of F
 WITHIN+RESIDUAL           60.00       16      3.75
 TYP                       24.00        1     24.00       6.40      .022
 VOL(1) W TYP(1)            3.00        1      3.00        .80      .384
 VOL(2) W TYP(1)            1.50        1      1.50        .40      .536
 VOL(3) W TYP(1)            1.50        1      1.50        .40      .536
 VOL(1) W TYP(2)           75.00        1     75.00      20.00      .000
 VOL(2) W TYP(2)           13.50        1     13.50       3.60      .076
 VOL(3) W TYP(2)            1.50        1      1.50        .40      .536
```

GLM uses /LMATRIX for simple effects contrasts with the coefficients for one level of the other factor set to 0, similar to earlier analysis for simple effects. The commands are illustrated below; as before, coefficients for main effects are included as well as those for simple effects. The results below agree with previous calculations. The first example produces a summary table for each contrast by including separate /LMATRIX options. No overall simple effect is presented. The second example produces the overall simple effect using /LMATRIX by including k - 1 orthogonal contrasts separated by semi-colons. Each contrast is tested for significance before the overall summary table.

```
GLM mis BY typ vol /LMATRIX vol -3 -1  1  3 vol BY typ  0 0 0 0 -3 -1  1  3
                   /LMATRIX vol  1 -1 -1  1 vol BY typ  0 0 0 0  1 -1 -1  1
                   /LMATRIX vol -1  3 -3  1 vol BY typ  0 0 0 0 -1  3 -3  1.
...
custom hypothesis tests #1
 l1       contrast estimate          22.000    SS = 3×22²/20 = 72.6
          std. error                 5.000     t = 22.0/5.0 = 4.4 = √F
          sig.                       .000

 source   sum of squares df mean square f      sig.
 contrast 72.600         1  72.600      19.360 .000
 error    60.000         16 3.750

custom hypothesis tests #2
 l1       contrast estimate          -4.000
          std. error                 2.236
          sig.                       .093

 source   sum of squares df mean square f     sig.
 contrast 12.000         1  12.000      3.200 .093
 error    60.000         16 3.750

Custom Hypothesis Tests #3
 L1       Contrast Estimate          -6.000
          Std. Error                 5.000
          Sig.                       .248
```

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	5.400	1	5.400	1.440	.248
Error	60.000	16	3.750		

```
GLM mis BY typ vol /LMATRIX vol -3 -1  1  3 vol BY typ  0 0 0 0  -3 -1  1  3;
                            vol  1 -1 -1  1 vol BY typ  0 0 0 0   1 -1 -1  1;
                            vol -1  3 -3  1 vol BY typ  0 0 0 0  -1  3 -3  1.
...
Custom Hypothesis Tests
 L1       Contrast Estimate          22.000          t = 22.0/5.0 = 4.4
          Std. Error                 5.000           t² = 4.4² = 19.36 = F_L1
          Sig.                       .000

 L2       Contrast Estimate          -4.000
          Std. Error                 2.236
          Sig.                       .093

 L3       Contrast Estimate          -6.000
          Std. Error                 5.000
          Sig.                       .248
```

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	**90.000**	**3**	**30.000**	**8.000**	**.002**
Error	60.000	16	3.750		

A simple effects analysis, perhaps followed by post hoc or planned comparisons of the simple effects, is one way to follow-up an interaction that was significant or predicted prior to the study. However, simple effects analyses sum SSs for a main effect and an interaction, which is partitioned into simple effects. The analysis does not involve pure interaction. An alternative approach is partitioning the interaction itself.

Partitioning the Interaction

It is possible to define contrast coefficients that partition $SS_{Interaction}$ into components associated with particular patterns in the data. Although simple to analyze, the results can be a challenge to interpret. The default factorial ANOVA for errors as a function of distraction by two types of sound (Noise or Speech) at four volumes showed a marginally significant interaction with df = 3. The SS and df for the interaction can be partitioned into three single df effects, as done previously for main effects. Here is the initial default analysis. Partitioning the interaction divides SS = 36.00 into three single df effects.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
typ	24.000	1	24.000	6.400	.022
vol	60.000	3	20.000	5.333	.010
typ * vol	36.000	3	12.000	3.200	.052
Error	60.000	16	3.750		

Partitioning the 36.00 units of variability in the T×V interaction takes two steps. First, contrast coefficients are created for the two main effects using the eight cell means rather than row and column means. The *typ* factor requires a single contrast to compare the two groups; this contrast is redundant with the main effect of *typ*, $df_{Type} = 1$. The *vol* factor with df = 3 requires three contrasts, as shown previously.

Polynomial contrasts are appropriate given the ordered levels of the volume factor.

Second, each contrast coefficient for one factor is multiplied times each contrast coefficient for the other factor to create (A-1)×(B-1) additional contrasts. These are interaction contrasts. The present study creates (2-1)×(4-1) = 3 new contrasts, as shown below. To illustrate this second step, the T×Vlin contrasts are: -1×-3 = +3, -1×-1 = +1, ..., +1×+1 = +1, and +1×+3 = +3. If Type had more than two levels then there would be two or more contrast coefficients for Type. Each contrast for Type would be multiplied times each contrast for Volume. See Appendix 6-1 for an example of two factors with $df_{Numerator}>1$.

	Noise				Speech					
	1	2	3	4	1	2	3	4		
$\overline{y}_{ab}$	4.0	3.0	5.0	4.0	2.0	5.0	9.0	8.0	L	SS
Main Effects										
Type -1	-1	-1	-1	+1	+1	+1	+1	8.0	$24.0 = 3\times 8^2/8 = SS_{Type}$ Main Effect	
V_{lin}	-3	-1	+1	+3	-3	-1	+1	+3		
V_{qua}	+1	-1	-1	+1	+1	-1	-1	+1		
V_{cub}	-1	+3	-3	+1	-1	+3	-3	+1		
Interaction										
TxV_{lin}	**+3**	**+1**	**-1**	**-3**	**-3**	**-1**	**+1**	**+3**	**20.0**	**$30.0 = 3\times 20^2/40$**
TxV_{qua}	-1	+1	+1	-1	+1	-1	-1	+1	-4.0	$6.0 = 3\times -4^2/8$
TxV_{cub}	+1	-3	+3	-1	-1	+3	-3	+1	0.0	$0.0 = 3\times 0^2/40$
									$\Sigma =$	$36.0 = SS_{TxV}$

Given the multiplication by -1 and +1, the interaction contrasts have opposite signs for the two levels of *typ*. The T×Vlin coefficients are +3, +1, -1, and -3 for the Noise condition and -3, -1, +1, and +3 for the Speech condition, a linear decrease for Noise and linear increase for Speech. These coefficients test for **opposite** linear patterns for the *vol* effect at each level *typ*; that is, a cross-over effect. The $T\times V_{qua}$ coefficients are an inverted u-shape for Noise and a u-shape for Speech, again opposite patterns for the two stimulus types. The specific positive or negative values are irrelevant as long as they reverse for the two levels of *typ*, because it is the strength of the correlation between these coefficients and the data that matters, not the sign. Also, the contrast L is squared to produce SS_L so the sign is irrelevant to the calculation, although not for the interpretation.

Once the interaction contrasts have been created, the analysis proceeds as for any contrast: $L = \Sigma c_j \overline{y}_j$ and $SS = n_j L^2 / \Sigma c_j^2$. These calculations are shown above. Most of the interaction variability falls on the $T\times V_{lin}$ contrast, which produces $F = 30.0/3.75 = 8.0$. This contrast is significant, whereas the others are not. Indeed SS = 0 for the $T\times V_{cub}$ component of the interaction. The SSs for the three contrasts sum to $SS_{T\times V}$, which has been partitioned into df = 1 effects with most variability loading on a single contrast, the linear.

This partitioning was produced in an earlier MANOVA analysis of main effects because the

SINGLEDF option partitions both main effects and interactions with df>1. The analysis is reproduced below. The 1st, 2nd, and 3rd parameters for typ BY vol agree with our calculations. The $T \times V_{lin}$ effect is highly significant, $p = .012$, indicating that the linear relationship between volume and mistakes is significantly different for Noise and Speech. As shown previously for main effects, the mean of the three *F*s equals the omnibus *F*; that is, $(8.0+1.6+0.0)/3 = 3.2 = F_{TxV}$, showing again that a specific contrast can be more significant than an omnibus effect. The t-tests and associated coefficients are not shown, but would be based on normalized coefficients.

```
MANOVA mis BY typ(1 2) vol(1 4) /CONTRAST(vol) = POLY /PRINT = SIGNI(SINGLEDF).
 Source of Variation          SS      DF        MS          F  Sig of F
 WITHIN CELLS              60.00      16      3.75
 typ                       24.00       1     24.00       6.40      .022
 vol                       60.00       3     20.00       5.33      .010
   1ST Parameter           43.20       1     43.20      11.52      .004
   2ND Parameter            6.00       1      6.00       1.60      .224
   3RD Parameter           10.80       1     10.80       2.88      .109
 typ BY vol                36.00       3     12.00       3.20      .052
   1ST Parameter           30.00       1     30.00       8.00      .012
   2ND Parameter            6.00       1      6.00       1.60      .224
   3RD Parameter             .00       1       .00        .00     1.000

 (Model)                  120.00       7     17.14       4.57      .006
 (Total)                  180.00      23      7.83
...
```

The corresponding analysis with GLM uses the /LMATRIX option and the contrast coefficients created above. Hence, GLM entails prior work to generate coefficients for the analysis. Unlike /LMATRIX for main and simple effects, vol BY typ appears without the *vol* main effect coefficients preceding it.

```
GLM mis BY typ vol /LMATRIX vol BY typ  +3  1 –1 –3  –3 –1  1  3
                   /LMATRIX vol BY typ  –1 +1 +1 –1  +1 –1 –1 +1
                   /LMATRIX vol BY typ  +1 –3 +3 –1  –1 +3 –3 +1.
...
Custom Hypothesis Tests #1
 L1       Contrast Estimate          20.000
          Std. Error                 7.071
          Sig.                       .012

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 30.000         1  30.000      8.000 .012
 Error    60.000         16 3.750

Custom Hypothesis Tests #2
 L1       Contrast Estimate          –4.000
          Std. Error                 3.162
          Sig.                       .224
 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 6.000          1  6.000       1.600 .224
 Error    60.000         16 3.750

Custom Hypothesis Tests #3
 L1       Contrast Estimate          .000
          Std. Error                 7.071
          Sig.                       1.000
```

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	.000	1	.000	.000	1.000
Error	60.000	16	3.750		

But what does it mean that $T \times V_{lin}$ is significant and accounts for most of the interaction variability? In its simplest terms, it means that contrast coefficients with reverse linear patterns correlate well with the actual data. We show this explicitly shortly, but can get some sense of the fit by returning to an earlier plot of the cell means with main effects removed; these means represent pure interaction. Figure 6-1 shows the plot along with labels for the contrast coefficients corresponding to each cell mean. The distinct pattern of decreasing mistakes with volume for the Noise condition, and increasing mistakes for the Speech condition is mirrored in the corresponding contrast coefficients. That is, there is a strong positive correlation between the linear contrast coefficients and cell means with main effects removed. If the signs of the contrast coefficients were reversed, the strong negative correlation would capture the same amount of variability in the interaction.

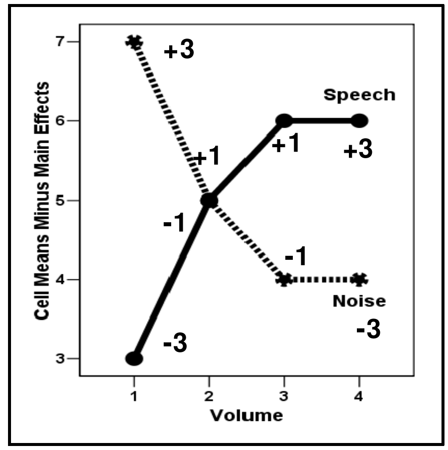


Figure 6-1. Interaction deviations and contrast coefficients

The interaction cell means also illustrate why the $T \times V_{qua}$ effect captured some of the variability in the means. The quadratic coefficients, -1 +1 +1 -1 and +1 -1 -1 +1, capture opposite U-shaped patterns. The two lines above have opposite curves after linear effects are removed. Noise shows a U-shaped pattern, whereas Speech shows an inverted U-shaped pattern. This accounts for 6.0 units of variability in the interaction and is not significant. Once the linear and quadratic components are removed, the Cubic coefficients, -1 +3 -3 +1 & 1 -3 3 -1 look for a zig-zag pattern in the data (i.e., up-down-up or down-up-down). In the present data, no zig-zag pattern remains in the means after main effects and linear and quadratic components of the interaction are removed.

It is important to remember that main effects are removed in Figure 6-1. In fact the actual change in mistakes with increasing volume for the Noise condition is minimal; the cell means are quite flat, as shown in Figure 6-2. Fitting separate straight lines to the Noise and Speech data, as done for regression, the slope for Noise would be small whereas the slope for Speech would be relatively large. This difference in slopes is captured by the $T \times V_{lin}$ coefficients.

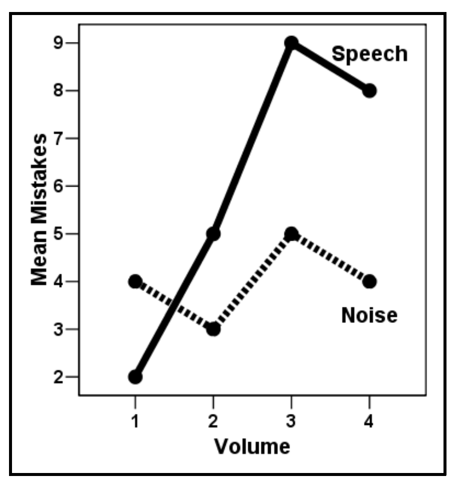


Figure 6-2. Plot of cell means

It is possible to “see” interaction effects in the original means. The difference in linear effects is clear in Figure 6-2. The means for noise are all similar across levels of Volume while those for Speech increase. As well the Speech condition shows an inverted U-shaped pattern about the linear effect, whereas the Noise condition shows little or no U-shaped pattern. That the cubic zig-zag patterns are identical is more difficult to see.

Thinking about interactions in terms of regression provides another way to conceptualize the linear component of the volume by type interaction. Figure 6-3 shows a scattergram of mistakes plotted as a function of volume separately for the Noise condition (stars and dashed line) and the Speech condition (boxes and solid line). The slopes for the two best-fit lines (i.e., linear relationships) are markedly different, with volume accounting for much variability in the Speech condition (r^2 = .605) and very little in the Noise condition (r^2 = .017). In regression, this represents the difference between slopes for best-fit lines, which is captured in ANOVA with the interaction partitioned into linear, quadratic, and cubic effects. We explore this further when examining a regression approach to these analyses.

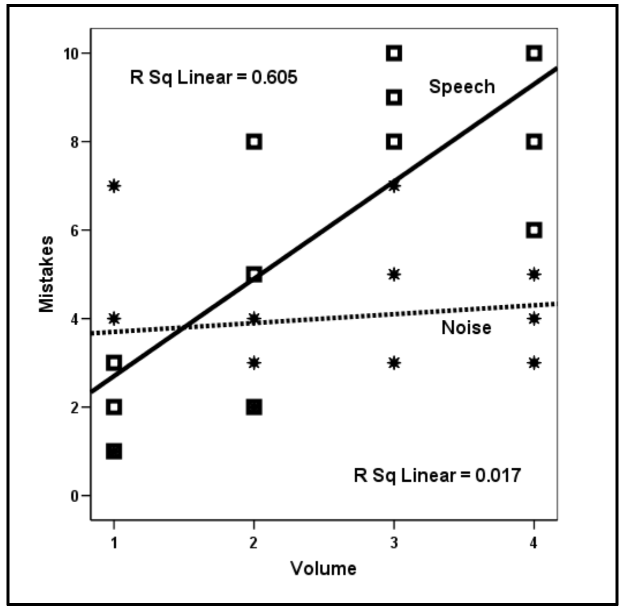


Figure 6-3. Separate regression lines for Noise and Speech

Before moving on to the regression analyses for factorial designs, however, one aspect of the GLM /LMATRIX option should be emphasized. As mentioned previously, the order of contrast coefficients *must* correspond to the order of the cell means in the GLM main command, and the order of cell means depends on the order that factors are listed after BY. The levels and means of the second factor after BY occur within the levels of the first factor and change more quickly. For BY *typ vol* cell means for *vol* are nested within levels of *typ*. Given BY *vol typ*, levels for *typ* are nested within levels of *vol* and contrast coefficients must be reordered to obtain correct results. One way to check that data and contrast coefficients align properly is to compare the order of cell means from the DESCRIPTIVE option and the order of contrast coefficients, as shown to the right of the cell means below, first for the *typ vol* order, and then for the *vol typ* order.

```
GLM mis BY typ vol /PRINT = DESCR /LMATRIX vol BY typ  +3 +1 −1 −3  −3 −1  1  3.

 typ    vol    Mean     Std. Deviation N              Cjs
 1.0000 1.0000 4.000000 3.0000000      3              +3
        2.0000 3.000000 1.0000000      3              +1
        3.0000 5.000000 2.0000000      3              −1
        4.0000 4.000000 1.0000000      3              −3
        Total  4.000000 1.8090681      12
 2.0000 1.0000 2.000000 1.0000000      3              −3
        2.0000 5.000000 3.0000000      3              −1
        3.0000 9.000000 1.0000000      3              +1
        4.0000 8.000000 2.0000000      3              +3
        Total  6.000000 3.3028913      12
 Total  1.0000 3.000000 2.2803509      6
        2.0000 4.000000 2.2803509      6
        3.0000 7.000000 2.6076810      6
        4.0000 6.000000 2.6076810      6
        Total  5.000000 2.7975144      24
...
Custom Hypothesis Tests
 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 30.000         1  30.000      8.000 .012
 Error    60.000         16 3.750

GLM mis BY vol typ /PRINT = DESCR /LMATRIX vol BY typ  +3 −3 +1 −1  −1 +1 −3 +3.

 vol    typ    Mean     Std. Deviation N                Cjs
 1.0000 1.0000 4.000000 3.0000000      3                +3
        2.0000 2.000000 1.0000000      3                −3
        Total  3.000000 2.2803509      6
 2.0000 1.0000 3.000000 1.0000000      3                +1
        2.0000 5.000000 3.0000000      3                −1
        Total  4.000000 2.2803509      6
 3.0000 1.0000 5.000000 2.0000000      3                −1
        2.0000 9.000000 1.0000000      3                +1
        Total  7.000000 2.6076810      6
 4.0000 1.0000 4.000000 1.0000000      3                −3
        2.0000 8.000000 2.0000000      3                +3
        Total  6.000000 2.6076810      6
 Total  1.0000 4.000000 1.8090681      12
        2.0000 6.000000 3.3028913      12
        Total  5.000000 2.7975144      24
...
Custom Hypothesis Tests
 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 30.000         1  30.000      8.000 .012
 Error    60.000         16 3.750
```

The first analysis above corresponds to how the cell means and contrast coefficients were arranged earlier. The second analysis corresponds to a different ordering of means and coefficients. In both cases, the coefficients on /LMATRIX correspond to the proper cell means in the descriptives.

Calculations with cells in different order

	Vol	1		2		3		4		
	Typ	N	S	N	S	N	S	N	S	
	$\bar{y}_{ab}$	4.0	2.0	3.0	5.0	5.0	9.0	4.0	8.0	
Main Effects										
TYP		−1	+1	−1	+1	−1	+1	−1	+1	
Vlin		−3	−3	−1	−1	+1	+1	+3	+3	
Interaction										L
TxVlin		+3	−3	+1	−1	−1	+1	−3	+3	20.0

Note the nonsense below when cell means and coefficients do not align properly with one another.

Coefficients in **WRONG** order

Vol	1		2		3		4	
Typ	N	S	N	S	N	S	N	S
$\bar{y}_{ab}$	4.0	2.0	3.0	5.0	5.0	9.0	4.0	8.0
Vlin	−3	−1	+1	+3	−3	−1	+1	+3

Regression Analyses for the Between-S Factorial

The single-factor study required k - 1 predictors to carry out analysis of variance by regression, where k equalled the number of levels of our factor. The present factorial study involves 8 cells, and requires 8 - 1 = 7 predictors. Partitioning the interaction involved 7 contrasts, one for *typ*, three for the main effect of *vol*, and three for the *typ* by *vol* interaction. To obtain the preceding analyses with REGRESSION, the 7 contrasts are used to create the 7 predictor variables with the following commands. The actual predictors are shown following the regression analysis.

```
RECODE typ (1 = −1) (2 = +1) INTO type.
RECODE vol (1 = −3) (2 = −1) (3 =  1) (4= 3) INTO vlin.
RECODE vol (1 =  1) (2 = −1) (3 = −1) (4= 1) INTO vqua.
RECODE vol (1 = −1) (2 =  3) (3 = −3) (4= 1) INTO vcub.
COMPUTE txvlin = type*vlin.
COMPUTE txvqua = type*vqua.
COMPUTE txvcub = type*vcub.
```

Regressing mistakes on these predictors shows the significance of each single df contrast (i.e., predictor), but does not provide information about the overall main effect of *vol* (*df* = 3) or the overall interaction (*df* = 3). These additional statistics can be obtained by entering the three predictors for the main effect OR the three predictors for the interaction after the other four predictors, and requesting CHANGE statistics for the strength and significance of the change in $SS_{Regression}$ when the final three predictors are added. F_{Change} will represent the effect of the three additional predictors (i.e., df = 3), either main effect of *vol* or interaction. The CHANGE statistic in the following analysis tests the overall interaction effect.

```
REGRESS /DESCR /STAT = DEFAU CHANGE /DEP = mis
  /ENTER type vlin vqua vcub /ENTER txvlin txvqua txvcub
  /SAVE PRED(prdmab) RESI(resmab).
```

	Mean	Std. Deviation	N	
mis	5.000	2.7975	24	***SS_{Total} = (24−1)×2.7975² = 180.0***
type	.00	1.022	24	
vlin	.00	2.284	24	
vqua	.00	1.022	24	
vcub	.00	2.284	24	***Ms = 0 for all predictors, contrasts***
txvlin	.00	2.284	24	
txvqua	.00	1.022	24	
txvcub	.00	2.284	24	

Correlations

	mis	type	vlin	vqua	vcub	txvlin	txvqua	***$r^2 \times SS_{Total}$***	
type	.365							***24.0***	***= SS_{Type}***
vlin	.490	.000						***43.2***	
vqua	−.183	.000	.000					***6.0***	***∑ = 60.0***
vcub	−.245	.000	.000	.000				***10.8***	***= SS_{Vol}***
txvlin	.408	.000	.000	.000	.000			***30.0***	
txvqua	−.183	.000	.000	.000	.000	.000		***6.0***	***∑ = 36.0***
txvcub	.000	.000	.000	.000	.000	.000	.000	***0.0***	***= SS_{TxV}***

The descriptive statistics are informative. First, the means for all seven predictors are 0, indicating they are contrasts (i.e., $\sum c_j = 0$). Second, the seven predictors are uncorrelated; that is, they are orthogonal ($\sum c_j c'_j = 0$). Finally, the correlation between each predictor and the dependent variable *mis* can be squared and multiplied by SS_{Total} to produce SSs that correspond to earlier calculations for each contrast. Summing the three contrasts for main effects of *vol* produces SS_{Vol} and summing the three contrasts for the interaction produces SS_{TxV}. In essence, contrast analysis tests the significance of the correlation between the data and a pattern represented by contrast coefficients. The simple correlations correspond to earlier calculations because predictors are orthogonal.

With respect to the summary table for Model 2 below, SS_{Total} is partitioned into what we can predict overall and residual: $SS_{Residual} = 60.0 = SS_{Error}$ from the analysis of variance, with df = N-p-1 = 24-7-1 = 16, as for SS_{Error}. Therefore, the error terms are the same for regression and analysis of variance, that is, $MS_{Residual} = MS_{Error}$. The overall F itself is of little interest as it aggregates all the effects; that is, $SS_{Regression} = 120.0 = SS_{Typ} + SS_{Vol} + SS_{TxV} = SS_{Model}$. This same value is SS_{Model} in ANOVAs.

Model	R	R Square	Change Statistics R Square Change	F Change	df1	df2	Sig.
1	.683(a)	.467	.467	4.156	4	19	.014
2	.816(b)	.667	.200	**3.200**	**3**	**16**	**.052**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	84.000	4	21.000	4.156	.014(a)
	Residual	96.000	19	5.053		
	Total	180.000	23			

2	Regression	120.000	7	17.143	4.571	.006(b)
	Residual	60.000	16	3.750		
	Total	180.000	23			

SS_{Change} = 120.0 − 84.0 = 36.0 = SS_{TxV}

$MS_{Residual} = MS_{Error}$ because the Model 2 regression equation produces cell means as predicted values; hence, residuals are deviations of observed values from the cell means, exactly how SS_{Error} was calculated. Predicted and residual values appear in the *prdmab* and *resmab* columns below.

The *t*s, *p*s, and *F*s (computed to the right) for the individual predictors correspond to earlier contrast analyses. The main effect of *typ* is significant, as are the linear components of the main effect of *vol* and of the *typ* by *vol* interaction. Other effects are not significant.

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	
...							
2	(Constant)	5.000	.395		12.649	.000	***t^2=F***
	type	1.000	.395	.365	2.530	.022	***6.40***
	vlin	**.600**	**.177**	**.490**	**3.394**	**.004**	***11.52***
	vqua	−.500	.395	−.183	−1.265	.224	***1.60***
	vcub	−.300	.177	−.245	−1.697	.109	***2.88***
	txvlin	**.500**	**.177**	**.408**	**2.828**	**.012**	***8.00***
	txvqua	−.500	.395	−.183	−1.265	.224	***1.60***
	txvcub	.000	.177	.000	.000	1.000	***0.00***

Here are the coefficients created earlier and the predicted and residual values from the regression.

```
FORMAT typ vol type TO txvcub (F2.0) mis prdmab resmab (F4.1).
LIST.
```

typ	vol	mis	type	vlin	vqua	vcub	t×vlin	t×vqua	t×vcub	prdmab	resmab
1	1	1.0	−1	−3	1	−1	3	−1	1	4.0	−3.0
1	1	4.0	−1	−3	1	−1	3	−1	1	4.0	.0
1	1	7.0	−1	−3	1	−1	3	−1	1	4.0	3.0
1	2	4.0	−1	−1	−1	3	1	1	−3	3.0	1.0
1	2	2.0	−1	−1	−1	3	1	1	−3	3.0	−1.0
1	2	3.0	−1	−1	−1	3	1	1	−3	3.0	.0
1	3	5.0	−1	1	−1	−3	−1	1	3	5.0	.0
1	3	7.0	−1	1	−1	−3	−1	1	3	5.0	2.0
1	3	3.0	−1	1	−1	−3	−1	1	3	5.0	−2.0
1	4	3.0	−1	3	1	1	−3	−1	−1	4.0	−1.0
1	4	5.0	−1	3	1	1	−3	−1	−1	4.0	1.0
1	4	4.0	−1	3	1	1	−3	−1	−1	4.0	.0
2	1	3.0	1	−3	1	−1	−3	1	−1	2.0	1.0
2	1	2.0	1	−3	1	−1	−3	1	−1	2.0	.0
2	1	1.0	1	−3	1	−1	−3	1	−1	2.0	−1.0
2	2	5.0	1	−1	−1	3	−1	−1	3	5.0	.0
2	2	2.0	1	−1	−1	3	−1	−1	3	5.0	−3.0
2	2	8.0	1	−1	−1	3	−1	−1	3	5.0	3.0
2	3	10.0	1	1	−1	−3	1	−1	−3	9.0	1.0
2	3	9.0	1	1	−1	−3	1	−1	−3	9.0	.0
2	3	8.0	1	1	−1	−3	1	−1	−3	9.0	−1.0
2	4	6.0	1	3	1	1	3	1	1	8.0	−2.0
2	4	10.0	1	3	1	1	3	1	1	8.0	2.0
2	4	8.0	1	3	1	1	3	1	1	8.0	.0

The default regression provides statistics for each single df effect, corresponding to the main effect of *typ* and the partitioning of the *vol* main effect and the *typ BY vol* interaction into linear, quadratic, and cubic components. To obtain information about effects with df > 1, either aggregate the df = 1 statistics or enter multiple predictors last and request change statistics, as shown here for the interaction.

SS_{Change} between Model 1 (only main effect predictors) and Model 2 (main + interaction predictors) produces SS_{TxV}. The test of significance for this change produced $F = 3.2$ and $p = .052$. These values correspond with those obtained earlier for the overall interaction effect.

We previously calculated the predicted cell means with no interaction by saving predicted scores from a GLM with only main effects; that is,

```
GLM mis BY typ vol /DESIGN typ vol
 /SAVE PRED(mnmain).
```

The same result can be produced by regressing mistakes on only the main effect predictors; that is,

```
REGRESS /DEP = mis /ENTER type TO vcub
  /SAVE PRED(prdmain).
```

The new variable, *prdmain*, contains cell means with the interaction removed, as shown in Figure 6-5. Deviations of *prdmab*, calculated above, from *prdmain* are the interaction deviations; that is, how far from no interaction are observed cell means. Squared and summed over all observations these deviations produce SS_{TxV}, as shown previously with GLM.

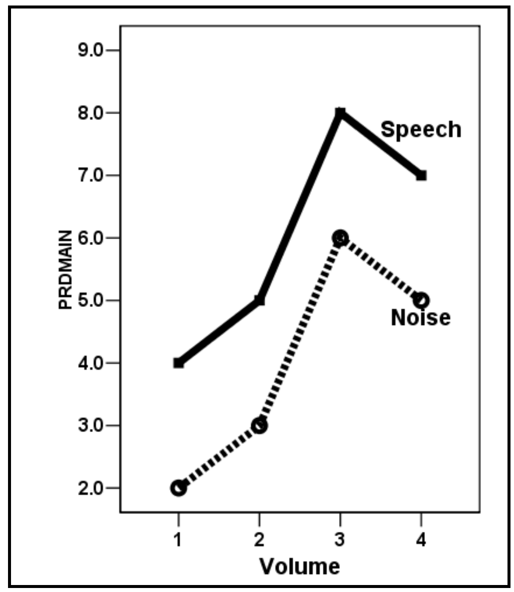


Figure 6-5. Predicted cell means from main effects.

Regression and the Simple Effects ANOVA

Earlier analyses of the simple effects of *vol* within levels of the *typ* factor again used 7 contrasts: one contrast for *typ*, three contrasts for *vol W typ(1)*, and three for *vol W typ(2)*. These seven contrasts are created below followed by a regression analysis that provides results for individual contrasts and for the overall simple effect of volume for the Speech condition in the CHANGE statistics.

The following commands generate the predictors or indicator variables. The *typ* factor is coded as before, but now linear, quadratic, and cubic predictors are created separately for the Noise and Speech conditions. The IFs below set the values of the predictors to 0 for the other *typ* condition; for example, the command IF typ = 2 vlinWt1 = 0 sets *vlinWt1* to 0 for the Speech condition. These new predictors are shown after the regression analysis.

```
RECODE typ (1 = -1) (2 = +1) INTO type.
RECODE vol (1 = -3) (2 = -1) (3 =  1) (4= 3) INTO vlinWt1.
IF typ = 2 vlinWt1 = 0.
```

```
RECODE vol (1 =  1) (2 = -1) (3 = -1) (4= 1) INTO vquaWt1.
IF typ = 2 vquaWt1 = 0.
RECODE vol (1 = -1) (2 =  3) (3 = -3) (4= 1) INTO vcubWt1.
IF typ = 2 vcubWt1 = 0.
RECODE vol (1 = -3) (2 = -1) (3 =  1) (4= 3) INTO vlinWt2.
IF typ = 1 vlinWt2 = 0.
RECODE vol (1 =  1) (2 = -1) (3 = -1) (4= 1) INTO vquaWt2.
IF typ = 1 vquaWt2 = 0.
RECODE vol (1 = -1) (2 =  3) (3 = -3) (4= 1) INTO vcubWt2.
IF typ = 1 vcubWt2 = 0.
```

The initial statistics show that the predictors are contrasts and orthogonal. Moreover, *r*s between predictors and *mis* reflect relationships between contrast coefficients and means. SSs are shown to the right for *vol W typ(2)*, separately for each contrast and then summed to obtain the overall simple effect for the Speech condition.

```
REGRESS /DESCR /STAT = DEFAU CHANGE /DEP = mis
  /ENTER type vlinWt1 vquaWt1 vcubWt1
  /ENTER vlinWt2 vquaWt2 vcubWt2
  /SAVE PRED(prdmab) RESI(resmab).
```

	Mean	Std. Deviation	N
mis	5.000	2.7975	24
type	.000000	1.0215078	24
vlinWt1	.000000	1.6151457	24
vquaWt1	.000000	.7223151	24
vcubWt1	.000000	1.6151457	24
vlinWt2	.000000	1.6151457	24
vquaWt2	.000000	.7223151	24
vcubWt2	.000000	1.6151457	24

	mis	type	vlinWt1	vquaWt1	vcubWt1	vlinWt2	vquaWt2	r^2 x SS_{Total}	
type	.365								
vlinWt1	.058	.000							
vquaWt1	.000	.000	.000						
vcubWt1	-.173	.000	.000	.000					
vlinWt2	.635	.000	.000	.000	.000			**72.6**	
vquaWt2	-.258	.000	.000	.000	.000	.000		**12.0**	**$\sum$=90 = SS_{VwT2}**
vcubWt2	-.173	.000	.000	.000	.000	.000	.000	**5.4**	

The summary table for Model 2 corresponds to that for the default factorial regression because predicted values equal the eight cell means. CHANGE statistics correspond to the overall simple effect of *vol* for the Speech condition, and individual predictors are planned contrasts for the simple effects.

Model	R	R Square	Change Statistics: R Square Change	F Change	df1	df2	Sig.
1	.408(a)	.167	.167	.950	4	19	.457
2	.816(b)	.667	.500	**8.000**	**3**	**16**	**.002**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30.000	4	7.500	.950	.457(a)
	Residual	150.000	19	7.895		
	Total	180.000	23			
2	Regression	120.000	7	17.143	4.571	.006(b)
	Residual	60.000	16	3.750		
	Total	180.000	23			

SS_{Change} = 90.0 = SS_{VwT2}

```
 Model            Unstandardized             Standardized    t       Sig.
                  Coefficients               Coefficients
                  B               Std. Error Beta
...
 2     (Constant) 5.000           .395                       12.649 .000
       type       1.000           .395       .365            2.530  .022
       vlinWt1    .100            .250       .058            .400   .694
       vquaWt1    .000            .559       .000            .000   1.000
       vcubWt1    -.300           .250       -.173           -1.200 .248
       vlinWt2    1.100           .250       .635            4.400  .000 F = 4.4² = 19.36
       vquaWt2    -1.000          .559       -.258           -1.789 .093
       vcubWt2    -.300           .250       -.173           -1.200 .248
```

The listing below shows predictors based on simple effect contrasts. They correspond to coefficients used previously to calculate SS by hand. The listing also shows that the predicted values equal the cell means and the residual values are deviations of the observed scores from the cell means, just how SS_{Error} was calculated using the ANOVA formula.

```
FORMAT typ vol type TO vcubWt2 (F2.0) mis prdmab resmab (F4.1).
LIST.
typ vol  mis type vlinWt1 vquaWt1 vcubWt1 vlinWt2 vquaWt2 vcubWt2 prdmab resmab
  1   1  1.0  -1      -3       1      -1       0       0       0     4.0   -3.0
  1   1  4.0  -1      -3       1      -1       0       0       0     4.0     .0
  1   1  7.0  -1      -3       1      -1       0       0       0     4.0    3.0
  1   2  4.0  -1      -1      -1       3       0       0       0     3.0    1.0
  1   2  2.0  -1      -1      -1       3       0       0       0     3.0   -1.0
  1   2  3.0  -1      -1      -1       3       0       0       0     3.0     .0
  1   3  5.0  -1       1      -1      -3       0       0       0     5.0     .0
  1   3  7.0  -1       1      -1      -3       0       0       0     5.0    2.0
  1   3  3.0  -1       1      -1      -3       0       0       0     5.0   -2.0
  1   4  3.0  -1       3       1       1       0       0       0     4.0   -1.0
  1   4  5.0  -1       3       1       1       0       0       0     4.0    1.0
  1   4  4.0  -1       3       1       1       0       0       0     4.0     .0
  2   1  3.0   1       0       0       0      -3       1      -1     2.0    1.0
  2   1  2.0   1       0       0       0      -3       1      -1     2.0     .0
  2   1  1.0   1       0       0       0      -3       1      -1     2.0   -1.0
  2   2  5.0   1       0       0       0      -1      -1       3     5.0     .0
  2   2  2.0   1       0       0       0      -1      -1       3     5.0   -3.0
  2   2  8.0   1       0       0       0      -1      -1       3     5.0    3.0
  2   3 10.0   1       0       0       0       1      -1      -3     9.0    1.0
  2   3  9.0   1       0       0       0       1      -1      -3     9.0     .0
  2   3  8.0   1       0       0       0       1      -1      -3     9.0   -1.0
  2   4  6.0   1       0       0       0       3       1       1     8.0   -2.0
  2   4 10.0   1       0       0       0       3       1       1     8.0    2.0
  2   4  8.0   1       0       0       0       3       1       1     8.0     .0
```

Comparison of Contrasts for Main, Simple, and Interaction Effects

It is informative to compare contrasts for main, simple, and interaction effects. The linear contrasts are shown below.

	Noise				Speech			
Vol =	1	2	3	4	1	2	3	4
$\overline{y}_{ab}$	4.0	3.0	5.0	4.0	2.0	5.0	9.0	8.0
Main Effect								
V_{lin}	−3	−1	+1	+3	−3	−1	+1	+3
Simple Effects								
V_{lin}Wt(1)	−3	−1	+1	+3	0	0	0	0
V_{lin}Wt(2)	0	0	0	0	−3	−1	+1	+3
Interaction								
TxV_{lin}	+3	+1	−1	−3	−3	−1	+1	+3

The main effect contrast tests for the *same* linear pattern in *vol* across both levels of *typ*. It tests whether the linear increase or decrease with volume is significant averaged across levels of type. The interaction contrast tests for the *opposite* linear pattern in *vol* at each level of *typ* (with main effects removed). It determines whether the slope for an increase or decrease in mistakes with volume at one level of the *typ* factor is significantly different from the slope for the other level of the *typ* factor. Finally, simple effects contrasts test the significance of *separate* linear patterns at each level of the *typ* factor. It determines whether the linear effect of volume is significant for the Noise type alone and whether the linear effect of volume is significant for the Speech type alone.

Although simple effects and partitioning were presented as two approaches to follow-up analyses for an interaction, the two methods can be used in conjunction with the omnibus *F* for the interaction. One possible scenario is first test the significance of the omnibus *F*, then the significance of the expected partitioning of the interaction, and finally the simple effects. This sequence provides increasingly specific information. In the present study, for example, the marginal omnibus *F* for the interaction and the graph suggest some difference in the effect of volume for Noise and Speech. Partitioning the interaction localizes the difference in the linear effects of the interaction. The simple effects analyses show that the linear effect of volume is not significant for Noise but is for Speech.

This completes discussion of the Between-S factorial design. We next consider designs involving one or more Within-Subject factors. For Within-S designs, numerators for the various effects are calculated and understood as for Between-S designs. However, different denominators are required to correctly test the null hypotheses. Understanding those denominators benefits from understanding interactions and simple-effects.

APPENDIX 6-1: INTERACTION CONTRASTS FOR FACTORIAL DESIGNS

The Type by Volume study was a 2×4 design, so partitioning the interaction involved one contrast for Type multiplied times three contrasts for Volume, resulting in three contrasts to partition the interaction. When both factors have more than two levels (i.e., are represented by more than a single contrast), then each contrast for one factor must be multiplied by each contrast for the other factor. Specifically, the (A-1) contrasts for factor A are multiplied times the (B-1) contrasts for factor B, which produces (A-1)(B-1) contrasts to partition the interaction. Recall that the df for the interaction is (A-1)(B-1) and each df requires a contrast. The example below shows the operations for a 3×3 design, which produces (3-1)(3-1) = 4 interaction contrasts. A 3×4 design would produce (3-1)(4-1) = 6 interaction contrasts, and so on.

		A1			A2			A3	
	B1	B2	B3	B1	B2	B3	B1	B2	B3
A_{linear}	−1	−1	−1	0	0	0	1	1	1
$A_{quadratic}$	−1	−1	−1	2	2	2	−1	−1	−1
B_{linear}	−1	0	1	−1	0	1	−1	0	1
$B_{quadratic}$	−1	2	−1	−1	2	−1	−1	2	−1
$A_{lin} \times B_{lin}$	1	0	−1	0	0	0	−1	0	1
$A_{lin} \times B_{qua}$	1	−2	1	0	0	0	−1	2	−1
$A_{qua} \times B_{lin}$	1	0	−1	−2	0	2	1	0	−1
$A_{qua} \times B_{qua}$	1	−2	1	−2	4	−2	1	−2	1

CHAPTER 7 - SINGLE-FACTOR WITHIN-S ANOVA

Previous analyses involved Between-S factors in which scores at one level of a factor are uncorrelated with scores at other levels of the factor. If scores are independent, it is reasonable to assume that all variability within levels of a factor is due to random variation or noise; hence the SSs and dfs within groups can be summed to create SS_{Error} and df_{Error}. This assumption is wrong when scores are correlated across conditions, which occurs when the same subjects are involved or scores correlate for other reasons (e.g., related subjects as in twins or animals from the same litter, subjects in different conditions matched on a relevant variable). Irrespective of how the expected correlation occurs, such factors are treated as Within-Subjects or Within-S factors. Between-S and Within-S analyses differ only with respect to error, the denominator for F. Conceptualization and previous calculations for numerators remain the same.

The following Within-S study involves four subjects who rated their agreement with controversial statements repeated for three trials. Ratings were obtained after each trial and scores appear below (higher scores = more agreement). Scores are expected to correlate across levels of the Trial factor if people with a certain level of belief on one trial are likely to have a similar level of belief on other trials as well. For example, subject 1 below had the lowest levels of belief for all three trials. Calculations are shown for SS_{Total} and SS_{Trial}. The calculation of these values is the same as for a Between-S factor.

	Trial (A)				
Subject	*1*	*2*	*3*		
1.	2	6	10		
2.	6	5	13		
3.	9	8	13		
4.	11	13	12		
$\bar{y}_a$	7.0	8.0	12.0	$\bar{y}_G$ =	9.0
$\bar{y}_a - \bar{y}_G$	−2.0	−1.0	+3.0	SS_{Total} =	146.0
SS_a	46.0	38.0	6.0		

For comparison purposes, the Between-S analysis is shown below. This is ***not*** a proper analysis for a Within-S design.

SS_{Total} = 146.0 = $SS_A + SS_{Error} = \sum\sum(y-\bar{y}_G)^2$

SS_A = 56.0 = $(-2^2 + -1^2 + 3^2)$ $\quad$ df_A = 3-1 = 2 $\quad$ MS_A = 28.0

SS_{Error} = 90.0 = $\sum SS_a$ = 46.0 + 38.0 + 6.0 = $SS_{Total} - SS_A$ $\quad$ df_{Error} = 12-3 = 9 $\quad$ MS_{Error} = 10.0

$F_{Between\text{-}S} = MS_A/MS_{Error}$ = 28.0 / 10.0 = 2.8 $\quad$ df = 2, 9 $\quad$ Between-S Analysis

The Between-S analysis is inappropriate because it assumes variability within-groups is entirely due to random error that can be aggregated (pooled) to create the error term for the *F*. In the Within-S design, however, some variability within groups is associated with systematic individual differences, and this shared variation should be removed from the denominator along with its degrees of freedom. Shared variability is removed in a paired-difference t-test, for example, by subtracting the two scores to obtain a single difference score. Within-S studies with more than two levels require an alternative approach, but it is equivalent to a paired-difference t-test when there are only two levels to the factor.

The Within-S approach treats Subjects as a second factor in the study, with each subject a level of the factor. Although we have not previously made the connection between error in the Between-S design and simple effects, SS_{Error} for the Between-S design is the sum of the simple effects of the Subjects factor within levels of A, that is, $SS_{Error} = SS_{SwA} = SS_{Sw1} + SS_{Sw2} + SS_{Sw3} = 46.0 + 38.0 + 6.0 = 90.0$. The Subject factor is nested within levels of factor A because scores in each group are not related. Recall from the factorial analysis that a simple effect is a combination of a main effect and an interaction. In this Between-S design, $SS_{SwTrial}$ combines the main effect of Subjects and the Trial by Subjects interaction. In the Within-S design, the Subject factor is said to be crossed with or orthogonal to levels of factor A as in a factorial ANOVA. Because scores in each group are related, the single-factor Within-S design is essentially a factorial design with A as one factor and Subjects (S) as a second factor. Each observation is analogous to a cell mean defined by levels of the two factors, with one observation for each cell.

Conceptualizing the Within-S design as a factorial study suggests that variability due to Subjects can be removed by calculating subject means averaged across treatments, and then computing $SS_{Subjects}$ (SS_S) just as SS_B was calculated in the factorial design. Calculations are shown below.

$SS_{Total} = SS_A + SS_{Subjects} + SS_{Error}$ $\quad\quad$ $SS_{Error} = SS_{Total} - SS_A - SS_S = SS_{AxS}$

	Trial (A)					
Subject	1	2	3	$\bar{y}_s$	$\bar{y}_s - \bar{y}_G$	n_s
1.	2	6	10	6.0	-3.0	3
2.	6	5	13	8.0	-1.0	3
3.	9	8	13	10.0	+1.0	3
4.	11	13	12	12.0	+3.0	3
$\bar{y}_a$	7.0	8.0	12.0	$\bar{y}_G = 9.0$		N = 12
n_a	4	4	4			
$\bar{y}_a - \bar{y}_G$	-2.0	-1.0	+3.0	SS_{Total} = 146.0		

$SS_S = 60.0 = 3(-3^2 + -1^2 + 1^2 + 3^2) = 60.0$ $\quad$ $df_S = 4 - 1 = 3$

$SS_A = 56.0 = 4(-2^2 + -1^2 +3^2) = 56.0$ $\quad$ $df_A = 3 - 1 = 2$ $\quad$ $MS_A = 56.0/2 = 28.0$

$SS_{Error} = 30.0 = 146.0 -56.0 -60.0 = 30.0$ $\quad$ $df_{Error} = 11 - 2 - 3 = 6$ $\quad$ $MS_{Error} = 30.0/6 = 5.0$

$F_{Within\text{-}S} = 28.0 / 5.0 = 5.6$ $\quad$ df = 2, 6 $\quad$ vs. $\quad$ $F_{Between\text{-}S} = 28.0 / 10.0 = 2.8$ $\quad$ df = 2, 9

Individual observations at each combination of factor A and factor S include an A×S interaction effect with $n_{as} = 1$ (i.e., 1 observation per cell). Once variability due to the main effects of A and S are removed, any remaining variability is due to the A×S interaction. The following calculations use our earlier formula for

SS_{AxB}. If scores are correlated, as in this example, then SS_{AxS} will be smaller than SS_{SwA} because SS_S is subtracted from SS_{Total}, leading to a larger value for F, albeit with fewer degrees of freedom. In general, this leads to greater likelihood of rejecting H_0: $\mu_1 = \mu_2 = ... \mu_k$ with a Within-S design. Within-S designs are more sensitive to differences among means.

Expected Values if no A×S interaction			Observed - Expected			
4	5	9	- 2	+1	+1	
6	7	11	0	- 2	+2	$1\times\Sigma^2 = 30.0 = SS_{Error}$
8	9	13	+1	- 1	0	
10	11	15	+1	+2	- 3	

e.g., 9.0 + -3.0 + -2.0 = 4 in top-left cell

SPSS Analyses for the Within-S ANOVA

To show that a Within-S analysis corresponds to a factorial design with A and S as factors, we first analyze the data in Between-S format, that is, with variables that represent the levels of A and S. Although the proper analysis can be done with the data organized as for a Between-S factorial like this, SPSS provides better ways to analyze Within-S factors as shown later. Most importantly, the Within-S approach in SPSS determines the appropriate error terms for the omnibus ANOVA and follow-up analyses, whereas errors often must be specified by users when data is in Between-S format. Remember that Within-S designs require different error terms than Between-S designs. Numerators remain the same.

```
*Data entered in Between-S format; one observation per cell.
DATA LIST FREE / subj trial agree.
BEGIN DATA
1 1 2   1 2 6   1 3 10       2 1 6   2 2 5   2 3 13
3 1 9   3 2 8   3 3 13       4 1 11  4 2 13  4 3 12
END DATA.

*Between-S ANOVA (Incorrect Error for Within-S study).
MANOVA agree BY trial(1 3) /PRINT = CELL /DESIGN trial.
      FACTOR           CODE                  Mean  Std. Dev.          N
  TRIAL                  1                  7.000      3.916          4
  TRIAL                  2                  8.000      3.559          4
  TRIAL                  3                 12.000      1.414          4
 For entire sample                          9.000      3.643         12

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              90.00       9     10.00
 TRIAL                     56.00       2     28.00      2.80      .113

 (Model)                   56.00       2     28.00      2.80      .113  ns
 (Total)                  146.00      11     13.27
```

The result matches earlier calculations for a Between-S design. This analysis is incorrect because it assumes that subject variability within groups is due to error. This gives, $SS_{Error} = 90.0$, which results in a failure to reject the H_0 that population means are equal ($p = .113$). Performing a full factorial design, as shown next, does not work because it leaves no source of variability for error. The main effects of T and S and the T×S interaction use up all the N-1 = 11 df. Note that N equals the number of observations in Within-S designs, not number of subjects. Each subject provides multiple observations. The /DESIGN option is used below to make explicit the default factorial analysis. The correct Within-S analysis extracts main effects of T and S, but uses the T×S interaction as the error; that is, $SS_{Error} = SS_{Trial \times Subject}$.

```
*Full-Factorial ANOVA (INCORRECT).
MANOVA agree BY trial(1 3) subj(1 4) /DESIGN trial subj trial BY subj.

 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 *   W A R N I N G   *  Too few degrees of freedom in RESIDUAL           *
 *                   *  error term for the following test(s) (DF = 0).  *
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 Source of Variation          SS      DF        MS         F  Sig of F
 RESIDUAL                    .00       0         .
 TRIAL                     56.00       2     28.00        .         .
 SUBJ                      60.00       3     20.00        .         .     Σdf=2+3+6=11
 TRIAL BY SUBJ             30.00       6      5.00        .         .

 (Model)                  146.00      11     13.27        .         .     df_Total=12-1=11
 (Total)                  146.00      11     13.27
```

To carry out the proper analysis with the Between-S format, SSs due to the main effects of Trial and Subjects are calculated, and the remaining variability, $SS_{T \times S}$, is error. That is, only main effects are entered on the /DESIGN statement. Removing SS_S reduces SS_{Error} relative to the Between-S analysis shown earlier. The main effect of Trial becomes significant given the smaller error term despite the fact that the numerator is unchanged and the loss of degrees of freedom. One important lesson about Within-S analyses is that they differ from Between-S analyses in the denominators, not the numerators. Note that $SS_{Residual} = SS_{A \times S}$.

```
*Main effects design for Within-S ANOVA in Between-S format (CORRECT).
MANOVA agree BY trial(1 3) subj(1 4) /DESIGN trial subj.

 Source of Variation          SS      DF        MS         F  Sig of F
 RESIDUAL                  30.00       6      5.00
 TRIAL                     56.00       2     28.00      5.60      .042
 SUBJ                      60.00       3     20.00      4.00      .070

 (Model)                  116.00       5     23.20      4.64      .044
 (Total)                  146.00      11     13.27
```

$SS_{A \times S}$ is used as error because we don't know why individual scores differ from what is expected on the basis of main effects of A (Trial in our example) and S. Subject 1 scored 2 units lower on trial 1 than expected given the main effects, but the reason is unknown. Subject 4 scored 2 units higher on trial 2 than expected, again for unknown reasons. The interaction is unexplained random variation or error.

Within-S factors are *not* normally analyzed with data in Between-S format although that approach is sometimes useful, such as with missing data. Normally, however, data is instead entered with multiple observations per case or subject, as for correlation analyses or the paired t-test. The ANOVA commands later "create" the within-S factor, as illustrated in the next few examples. Data entry is shown first. Three scores are entered per subject resulting in a data file with four rows (one per each subject) and three columns (one score for each trial, named t1, t2, and t3).

```
*Data entry for Within-S Format.
DATA LIST FREE / t1 t2 t3.
BEGIN DATA
2 6 10      6 5 13      9 8 13      11 13 12
END DATA.
```

To conduct the Within-S analysis, MANOVA and GLM use /WSF (i.e., Within-S Factor) to identify t1, t2, and t3 as levels of a Within-S factor. Considerable extra output is produced by both analyses, especially GLM. The extra output is italicized below and omitted in later printouts. Critical results are in bold.

```
*MANOVA (Note considerable extra output in italics).
MANOVA t1 t2 t3 /WSF = trial(3).

Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              60.00       3     20.00                          SSS dfS
 CONSTANT                 972.00       1    972.00     48.60      .006      ignore here

Tests involving 'TRIAL' Within-Subject Effect.
 Mauchly sphericity test, W =      .78667
 Chi-square approx. =              .47990 with 2 D. F.
 Significance =                      .787

 Greenhouse-Geisser Epsilon =      .82418
 Huynh-Feldt Epsilon =            1.00000
 Lower-bound Epsilon =             .50000

AVERAGED Tests of Significance that follow multivariate tests are equivalent to
univariate or split-plot or mixed-model approach to repeated measures.
Epsilons may be used to adjust d.f. for the AVERAGED results.

 EFFECT .. TRIAL
 Multivariate Tests of Significance (S = 1, M = 0, N = 0)

 Test Name          Value    Exact F Hypoth. DF   Error DF  Sig. of F
 Pillais          .77220    3.38983       2.00       2.00       .228
 Hotellings      3.38983    3.38983       2.00       2.00       .228
 Wilks            .22780    3.38983       2.00       2.00       .228
 Roys             .77220
 Note.. F statistics are exact.

Tests involving 'TRIAL' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              30.00       6      5.00                         SST
 TRIAL                     56.00       2     28.00      5.60      .042     SSTxS
```

The Between-Subjects Effect is variability due to Subjects, as calculated earlier. It serves as a denominator to test whether $\overline{y}_G$ differs significantly from 0, which is often of no interest but not always. The Within-S effect agrees with earlier calculations and with the proper analysis reported earlier with data in Between-S format. The same results are shown below for GLM, again with much extra output, including polynomial contrasts that are the default for Within-S factors. Much output consists of alternative analyses appropriate if the assumptions for the Within-S design are violated.

```
*GLM - Note MUCH extra output; focus on Sphericity Assumed lines.
GLM t1 t2 t3 /WSF = trial 3.
```

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
TRIAL	Pillai's Trace	.772	3.390(a)	2.000	2.000	.228
	Wilks' Lambda	.228	3.390(a)	2.000	2.000	.228
	Hotelling's Trace	3.390	3.390(a)	2.000	2.000	.228
	Roy's Largest Root	3.390	3.390(a)	2.000	2.000	.228

Within Subjects Design: TRIAL
Mauchly's Test of Sphericity(b)
Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon(a) Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TRIAL	.787	.480	2	.787	.824	1.000	.500

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

Tests of Within-Subjects Effects
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TRIAL	**Sphericity Assumed**	**56.000**	**2**	**28.000**	**5.600**	**.042**
	Greenhouse-Geisser	*56.000*	*1.648*	*33.973*	*5.600*	*.057*
	Huynh-Feldt	*56.000*	*2.000*	*28.000*	*5.600*	*.042*
	Lower-bound	*56.000*	*1.000*	*56.000*	*5.600*	*.099*
Error(TRIAL)	**Sphericity Assumed**	**30.000**	**6**	**5.000**		
	Greenhouse-Geisser	*30.000*	*4.945*	*6.067*		
	Huynh-Feldt	*30.000*	*6.000*	*5.000*		
	Lower-bound	*30.000*	*3.000*	*10.000*		

Source	TRIAL	Type III Sum of Squares	df	Mean Square	F	Sig.
TRIAL	Linear	50.000	1	50.000	10.000	.051
	Quadratic	6.000	1	6.000	1.200	.353
Error(TRIAL)	Linear	15.000	3	5.000		
	Quadratic	15.000	3	5.000		

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	972.000	1	972.000	48.600	.006
Error	**60.000**	**3**	**20.000**		

For a Within-S design, GLM can be accessed from menus with: *Analyze* | *General Linear Model* | *Repeated Measures*, which brings up the menu in Figure 7-1. Click in the *Within-Subject Factor Name* box to assign a name for the factor and in the *Number of Levels* box for the number of levels. Trial and 3 would be entered for our study. Click *Add* followed by *Define* to obtain the screen in Figure 7-2.

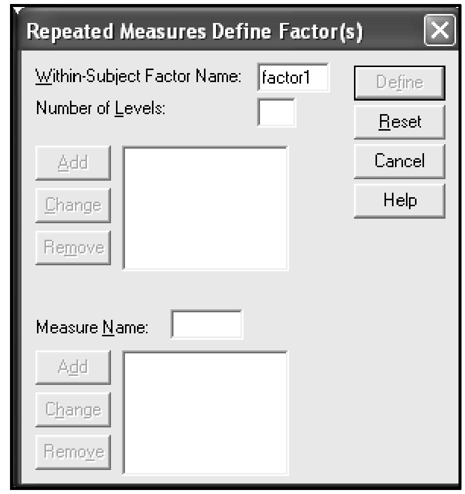


Figure 7-1. First Within-S menu

Figure 7-2 shows an empty frame for the Trial factor, with three slots for scores. The t1, t2, and t3 scores can be entered by selecting and moving them into the frame, either one at a time or all three at once *if* scores are in the same order as their slots in the frame.

Once the Within-S factor has been filled with variables, Between-S factors (see Chapter 9) or additional aspects of the design can be requested. Click *Ok* to run the analysis. GLM produces the following syntax, which includes default values for several GLM options.

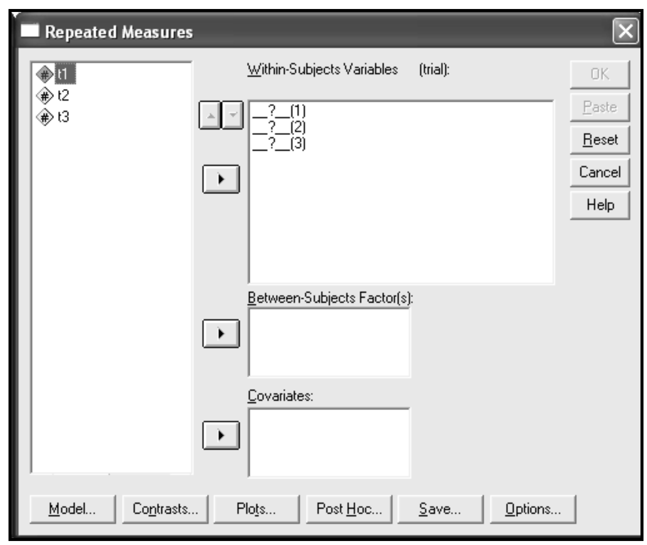


Figure 7-2. Second Within-S menu

```
GLM  t1 t2 t3
  /WSFACTOR = trial 3 Polynomial
  /METHOD = SSTYPE(3)
  /CRITERIA = ALPHA(.05)
  /WSDESIGN = trial.
```

The */METHOD* and */CRITERIA* options specify default values, as does the */WSDESIGN* option, which requests analysis for the main effect of trial. The /WSF option specifies Polynomial contrasts, the default contrast for Within-S factors.

Whether MANOVA or GLM is used, the conclusion from the Within-S analysis is to reject H_0: $\mu_1 = \mu_2 = \mu_3$, and accept the alternative that one or more equality is false. This vague conclusion requires follow-up analyses. Before turning to post hoc and planned analyses for this study, note again that the critical difference between the Between-S and Within-S analyses is the error term. The numerator for testing the null hypothesis is the same in both analyses, but different error terms are used. This is also also true for follow-up analyses; same calculations for numerators of follow-up analyses, but unique denominators.

Planned Contrasts

Partitioning the numerator SS_A for planned contrasts is identical to the Between-S design, as illustrated below with integer contrast coefficients. With normalized coefficients L_{linear}=-.7071 × 7.0 + 0 × 8.0 + .7071 × 12=3.5355. Note that $-.7071^2+0^2+.7071^2=1.0$ and $SS_L=4\times3.355^2/1=50.00$. But for Within-S designs, the denominator (i.e., $SS_{A\times S}$ = 30.0) is also partitioned to produce a unique Error term for each contrast. Polynomial contrasts are appropriate for this study and are the default contrasts for Within-S factors in GLM.

	1	2	3	L	SS	
$\overline{y}_a$	7.0	8.0	12.0	L	SS	
Linear	−1	0	+1	5.0	50.0	$SS_{Linear} = 4\times5^2/2$
Quadratic	−1	+2	−1	−3.0	6.0	
					Σ 56.0 = SS_A = SS_{Trial}	

For these contrasts, error is partitioned into linear and quadratic components. Because SS_{Error} is the A×S interaction, the error is partitioned by computing contrasts for relevant components of the A×S interaction. Note the relationship in the following SPSS analysis between the sum of the denominators for each contrast and the overall A×S error term

The SPSS commands to perform planned contrasts for Within-S factors are identical or similar to those for Between-S factors, but the output is quite different. One change for MANOVA is that single df effects for Within-S factors use the SIGNIF(UNIVARIATE) option; that is, different keywords are used for Within-S and Between-S factors to get single df tests. Also, the warning below from MANOVA informs users that MANOVA automatically created orthogonal and normalized contrasts, which could mean that aspects of analyses may not correspond to the requested contrasts unless they are orthogonal.

```
MANOVA t1 t2 t3 /WSF = trial(3) /PRINT = SIGNIF(UNIVARIATE)
  /CONTR(trial) = SPECIAL(1 1 1 -1 0 +1  -1 2 -1).

>Warning # 12252 in column 19.  Text: SPECIAL
>Special contrasts were requested for a WSFACTOR.  MANOVA automatically
>orthonormalizes contrast matrices for WSFACTORS.  If the special contrasts
>that were requested are nonorthogonal, the contrasts actually fitted are
>not the contrasts requested.  See the transformation matrix for the actual
>contrasts fitted.  Use TRANSFORM instead of WSFACTORS to produce
>nonorthogonal contrasts for within subjects factors.  Multivariate and
>averaged tests remain valid.

Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              60.00       3     20.00
 CONSTANT                 972.00       1    972.00     48.60      .006

 Estimates for T1 --- Individual univariate .9500 confidence intervals
 CONSTANT
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t Lower -95%  CL- Upper
        1   15.5884573    2.23607    6.97137      .00606    8.47229   22.70462
```

```
Tests involving 'TRIAL' Within-Subject Effect.
...
 Univariate F-tests with (1,3) D. F.
 Variable   Hypoth. SS   Error SS Hypoth. MS   Error MS          F  Sig. of F
 T2           50.00000   15.00000   50.00000    5.00000   10.00000       .051
 T3            6.00000   15.00000    6.00000    5.00000    1.20000       .353

 AVERAGED Tests of Significance for T using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              30.00       6      5.00
 TRIAL                     56.00       2     28.00      5.60      .042

 Estimates for T2 --- Individual univariate .9500 confidence intervals
 TRIAL
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t
        1   3.53553391    1.11803    3.16228     .05078           Normalized L

 Estimates for T3 --- Individual univariate .9500 confidence intervals
 TRIAL
  Parameter     Coeff.  Std. Err.    t-Value     Sig. t
        1   -1.2247449    1.11803   -1.09545     .35339
```

The critical output appears as a summary table above the overall Trial effect and in the equivalent t-tests that follow. The equivalence of *F* and *t* for corresponding contrasts is shown by equal *p* values and by $F = t^2$. SS_{Trial} has been partitioned into Linear and Quadratic components, which appear in the Hypoth. SS column. These agree with earlier calculations.

Similarly, $SS_{Error} = SS_{T\times S}$ has been partitioned into Linear and Quadratic components in the Error SS column. The Error SS for T2 (the linear contrast) agrees with later calculations. Although SS_{Error} has been divided equally here, generally the denominators for Within-S contrasts will not be identical, unlike Between-S contrasts. This can result in outcomes that appear paradoxical; for example, an effect with a smaller numerator could be significant and an effect with a larger numerator not, depending on their denominators.

Although the linear contrast captures most of the variability, resulting in a larger *F* than the omnibus *F*, the effect is only marginally significant here. This outcome occurs because $df_{Denominator} = 3$ for the contrast and df = 6 for the omnibus *F*. The loss of df is serious in this example given the extremely small sample size. With more subjects, the impact of a smaller df is more modest.

There are several ways to conceptualize the error terms for the contrasts. One way, albeit a clumsy one, is to actually partition the A×S interaction as done previously for the factorial A×B interaction. The method is awkward because it requires contrast coefficients for Subjects (which could have many levels) and for the factor, which are then multiplied to produce A×S contrast coefficients. The process is illustrated in Appendix 7-1.

$$\bar{L} = \sum c_j \times \bar{y}_j$$

$$SS_L = \frac{n_S \times (\bar{L} - 0)^2}{\sum c_j^2}$$

$$SS_{Error} = (n_S - 1) \times SD_L^2$$

Box 7-1. Within-S contrast scores

A simpler and more easily generalized way to obtain unique denominators for contrasts is to compute a contrast score for each subject, and determine the variability in those contrast scores, similar to difference scores for the paired-difference t-test. Calculations with normalized coefficients (i.e., integer coefficients divided by the square root of the sum of the integer coefficients squared) correspond better with all aspects of the SPSS output, although the ultimate statistics (*F*, *t*, *p*) are the same.

Box 7-1 shows the relevant formula. The following linear contrast uses normalized coefficients but the final F would be the same with integer coefficients.

				Linear (L)	
1.	2	6	10	5.6569	= -.7071×2.0+0×6.0+.7071×10.0
2.	6	5	13	4.9497	
3.	9	8	13	2.8284	$\bar{L}$ = 3.5355
4.	11	13	12	.7071	SD_L = 2.236

$SS_{Lin} = 4 \times (3.5355\text{-}0)^2 = 50.0$

$SS_{Error} = (4\text{-}1)2.236^2 = 15.00 = SS_{Lin\times Subj}$ df = 4 - 1 = 3

$F_{Linear} = (50.0/1) / (15.0/3) = 50.0 / 5.0 = 10.0$ $\sqrt{F} = 3.1623$

This procedure is a generalization of the paired-difference t-test, for which difference scores are contrast scores (i.e., D = $-1\times y_1 + 1\times y_2$ for each subject) to test the significance of H_0: $\mu_D = 0$ given the mean and standard deviation of the individual difference scores. The preceding contrast is identical to a paired-difference t-test because the Linear contrast compares condition 3 to condition 1. See below and Box 7-2 for relevant calculations.

T1	T3	D	$(D-\bar{D})^2$
2	10	8	9
6	13	7	4
9	13	4	1
11	12	1	16
		$\bar{D}$=5.0	SS_D= 30.0

$$\bar{D} = 5.0 \qquad SD_D = \sqrt{\frac{30}{4-1}} = 3.16228$$

$$t = \frac{5.0-0}{\frac{3.16228}{\sqrt{4}}} = \frac{5.0}{1.5811} = 3.1623 = \sqrt{F_{Linear}}$$

Box 7-2. Paired-Difference t-test

More on Planned Contrasts in SPSS

In addition to the UNIVARIATE option, Within-S contrasts in MANOVA can use /WSDESIGN (or the abbreviation /WSD) to request components of the Trial effect rather than the default omnibus effect. /WSDESIGN is analogous to /DESIGN, but for Within-S factors. MANOVA does not allow Within-S factors to be included on the /DESIGN option, or Between-S factors to be included on the /WSD option. A separate summary table and *t* is presented for each contrast.

```
MANOVA t1 t2 t3 /WSF = trial(3) /CONTR(trial) = POLY /WSD trial(1) trial(2) /DESIGN.
```

```
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           60.00       3     20.00
 CONSTANT                 972.00       1    972.00     48.60      .006

Tests involving 'TRIAL(1)' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           15.00       3      5.00
 TRIAL(1)                  50.00       1     50.00     10.00      .051
```

```
 Estimates for T2 --- Individual univariate .9500 confidence intervals
 TRIAL(1)
  Parameter           Coeff.         Std. Err.          t-Value           Sig. t
        1      3.5355339059         1.11803          3.16228           .05078

Tests involving 'TRIAL(2)' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           15.00       3      5.00
 TRIAL(2)                   6.00       1      6.00      1.20      .353

 Estimates for T3  --- Individual univariate .9500 confidence intervals
 TRIAL(2)
  Parameter           Coeff.         Std. Err.          t-Value           Sig. t
        1      1.2247448714         1.11803          1.09545           .35339
```

The following t-test shows the equivalence of the linear contrast (-1 0 +1) to a paired- difference T-test between trials one and three.

TTEST PAIR t1 t3.

```
...
                Mean           Std.      Std. Error    t   df  Sig.
                               Deviation Mean                  (2-tailed)
 Pair 1 T1 - T3 -5.000000      3.1622777 1.5811388   -3.162 3  .051
```

The corresponding GLM analysis appears below. The contrast is specified as part of the /WSF option, rather than as a separate command. The results agree with earlier calculations and the MANOVA analysis.

GLM t1 t2 t3 /WSFACTOR = trial 3 POLYNOMIAL.

```
Tests of Within-Subjects Effects
 Source                         Type III Sum of Squares df    Mean Square F     Sig.
 TRIAL        Sphericity Assumed 56.000                  2     28.000      5.600 .042
 Error(TIME)  Sphericity Assumed 30.000                  6     5.000

Tests of Within-Subjects Contrasts
 Source       TRIAL     Type III Sum of Squares df Mean Square F       Sig.
 TRIAL        Linear    50.000                  1  50.000      10.000 .051
              Quadratic 6.000                   1  6.000       1.200  .353
```

$\sum = 56.0 = SS_{Trial}$

```
 Error(TRIAL)Linear    15.000                  3  5.000
              Quadratic 15.000                  3  5.000
```

$\sum = 30.0 = SS_{Error}$

```
Tests of Between-Subjects Effects
 Source    Type III Sum of Squares df Mean Square F      Sig.
 Intercept 972.000                 1  972.000     48.600 .006
 Error     60.000                  3  20.000
```

Several other features of GLM for Within-S factors merit brief mention. First, if a SPECIAL set of contrasts is desired, then SPECIAL must include k 1s (as in MANOVA) followed by k-1 sets of k contrasts. This is illustrated below. Second, SSs for the two contrasts in the following analysis do *not* sum to SS_{Trial}. GLM used the integer coefficients provided in the command rather than normalized coefficients but did not adjust for the magnitude of the coefficients. The final *F*s do agree, so the conclusions are the same. To obtain

SSs corresponding to the partitioning of SS_{Trial}, normalized coefficients must be used. Third, the second analysis below illustrates the MMATRIX option, which is to Within-S factors what LMATRIX is to Between-S factors. However, the contrasts are specified somewhat differently. Normalized coefficients have been used below to produce SSs that correspond to partitioning SS_{Trial}. The output agrees with earlier results.

```
GLM t1 t2 t3 /WSF = trial 3 SPECIAL(1 1 1  -1 0 1  -1 2 -1).
...
Tests of Within-Subjects Effects
 Source                        Type III Sum of df    Mean Square F      Sig.
                               Squares
 trial         Sphericity      56.000          2     28.000      5.600 .042
               Assumed
 Error(trial)  Sphericity      30.000          6     5.000
               Assumed

Tests of Within-Subjects Contrasts
 Source        trial Type III Sum of df Mean Square F      Sig.
                     Squares
 trial         L1    100.000         1  100.000     10.000 .051     Fs correct
               L2    36.000          1  36.000      1.200  .353
                     ∑ = 136.0 ≠ SS_Trial

 Error(trial)  L1    30.000          3  10.000
               L2    90.000          3  30.000
...                  ∑ = 120.0 ≠ SS_Error

GLM t1 t2 t3 /WSF = trial 3 /MMATRIX t1 -.70711  t2 0  t3 .70711.
...
Custom Hypothesis Tests

 L1           Contrast Estimate          3.536        t=3.536/1.118=3.163=√F
              Std. Error                 1.118
              Sig.                       .051

 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 50.000         1  50.000      10.000 .051
 Error    15.000         3  5.000
```

As noted earlier, the parallel between the paired-difference t-test and the linear contrast provides some insight into the nature of planned contrasts for Within-S factors. In essence, subjects are given scores that represent how well their data corresponds to a predicted pattern, calculated by the following COMPUTES. The average of these scores is compared to an expected value of 0 given the null hypothesis of no such pattern in the data (i.e., $\mu_L = 0$), with the variability in contrast scores providing the proper error term.

```
COMP nrlin = -.70711*t1 +      0*t2 +.70711*t3.
COMP nrqua = -.40825*t1 + .81650*t2 -.40825*t3.

LIST t1 t2 t3 nrlin nrqua.
      T1        T2        T3    NRLIN    NRQUA
  2.0000    6.0000   10.0000   5.6569    .0000
  6.0000    5.0000   13.0000   4.9498  -3.6743
  9.0000    8.0000   13.0000   2.8284  -2.4495
 11.0000   13.0000   12.0000    .7071   1.2247
```

```
DESCR nrlin.
                      N Minimum Maximum Mean     Std. Deviation
 NRLIN                4 .7071   5.6569  3.535550 2.2360782
```

This gives all the information needed to carry out the planned linear contrast; that is,

$SS_{Contrast} = 4 \times (3.53555-0)^2 = 50.0$ df = 1 $MS_{Contrast} = 50.0$

$SS_{Error} = (4-1) \times 2.2360782^2 = 15.0$ df = 3 $MS_{Error} = 5.0 = 2.2360782^2$

$F = 50.0/5.0 = 10.0$ $\sqrt{10.0} = 3.162 = t$ (see below)

Alternatively, SPSS can perform *t* and *F* tests on the new scores. The single-sample t below determines whether the average linear score differs significantly from 0. The MANOVA below, without any Within-S or Between-S factors, illustrates one case when the test of whether $\overline{y}_G$ differs from 0 is meaningful. Because *nrlin* scores are contrast scores, their average should differ from 0 only by chance if there is no linear relationship in the data. GLM could also carry out this test as follows: GLM nrlin. Note the equivalence of the results: $t^2 = F$, p = .051.

```
TTEST TESTVALU 0 /VARI nrlin.
       N Mean     Std. Deviation Std. Error Mean
 NRLIN 4 3.535550 2.2360782      1.1180391                    2.2360782/√4 = 1.118

          t         df   Sig.       Mean
                        (2-tailed)  Difference
     NRLIN 3.162     3    .051       3.535550                 3.162²=10.0 = F

MANOVA nrlin.
 Tests of Significance for NRLIN using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              15.00       3      5.00
 CONSTANT                  50.00       1     50.00     10.00      .051
```

Post-Hoc Comparisons

Comparisons between pairs of means also require unique error terms, as just shown for the t3 vs t1 contrast. LSD comparisons are essentially paired t-tests without any adjustment, whereas various adjustments are used for other procedures (e.g., Bonferroni). The SPSS /POSTHOC procedure is not available for Within-S factors, so manual calculations may be necessary for some tests. GLM does allow a few options for post-hoc comparisons of Within-S factors (i.e., LSD, Bonferroni, Sidak) in the EMMEANS option. For the SNK and TUKEY procedures, one somewhat unorthodox procedure is to compute paired ts, multiply by $\sqrt{2}$ to obtain the corresponding qs, and obtain critical values using appropriate stretches.

The following GLM and TTEST analyses demonstrate that the LSD results produced by SPSS are simply paired-difference t-tests. The final GLM shows that p values for BONFERRONI are 3 times the LSD p values, as observed with Between-S factors.

```
GLM  t1 t2 t3  /WSFACTOR = trial 3  /EMMEANS = TABLES(trial) COMPARE ADJ(LSD).
...
Estimated Marginal Means
 (I)   (J)   Mean Difference Std.  Sig.(a) 95% Confidence Interval for
 trial trial (I-J)           Error         Difference(a)
                                           Lower Bound     Upper Bound
 1     2     -1.000          1.225 .474    -4.898          2.898
       3     -5.000          1.581 .051    -10.032         .032
 2     3     -4.000          1.871 .122    -9.954          1.954
```

```
TTEST PAIR t1 t2 t3.
...
                Paired Differences                         t      df   Sig.
                                                                          (2-tailed)
                Mean           Std.      Std. Error
                               Deviation Mean

 Pair 1 t1 - t2 -1.0000000     2.4494897 1.2247449         -.816       3  .474
 Pair 2 t1 - t3 -5.0000000     3.1622777 1.5811388         -3.162      3  .051
 Pair 3 t2 - t3 -4.0000000     3.7416574 1.8708287         -2.138      3  .122
```

```
GLM  t1 t2 t3  /WSFACTOR = trial 3  /EMMEANS = TABLES(trial) COMPARE ADJ(BONF).
...
Estimated Marginal Means  TRIAL
       Mean   Std.
              Error
 TRIAL
 1     7.000  1.958
 2     8.000  1.780
 3     12.000 .707

Pairwise Comparisons
                            Mean Difference Std.  Sig.(a)
                            (I-J)           Error
 (I) TRIAL (J) TRIAL
 1         2         -1.000          1.225 1.000
           3         -5.000          1.581 .152              p = 3 × .051
 2         3         -4.000          1.871 .366              p = 3 × .122
```

Conclusions

Although the preceding analyses can be done using ANOVA and data in Between-S format or even using REGRESSION, the analyses are messy because many indicator variables are required for large samples and users would need to specify the appropriate denominator for tests. Do remember, however, that the seeming “complexities” of error terms are in fact generalizations of prior principles (e.g., indicator variables for main and interaction effects, partitioning interactions, variability in difference scores, ...).

Given the analysis of a single Within-S factor, notably the selection of appropriate error terms for omnibus and follow-up analyses, chapter 8 considers factorial studies in which both factors are Within-S and chapter 9 analyses for factorials with both Within-S and Between-S factors. Again, differences between analyses for Within-S and Between-S factors will be the denominator or error terms for various effects. Conceptualization and calculation of the numerators involved in factorial effects, such as interaction and simple effects, are the same irrespective of the design.

Appendix 7-1: Partitioning the A×S Interaction

The denominator for a Within-S factor is the interaction between the factor (A) and subjects (S), which is partitioned along with the numerator for single df contrasts. As seen for factorial designs, partitioning an interaction involves multiplying the contrast coefficients for both factors. The coefficients used for the subject contrasts in a Within-S design are arbitrary, as long as they are orthogonal. The procedure is illustrated below for the linear contrast; note that this operation is not used in practice because a large number of contrasts would be required for any respectable number of subjects; for example, with 30 subjects and a factor with 3 levels, we would need 29 contrasts for subjects and 29 contrasts for the linear × subjects interaction. Instead of users having to generate contrasts, SPSS does the proper partitioning; however, the logic is essentially that involved in partitioning interactions. The chapter shows a "better" way to conceptualize the unique error terms, namely as variability in contrast scores for each subject.

		S1			S2			S3			S4		
	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3	
	2	6	10	6	5	13	9	8	13	11	13	12	
Lin	-1	0	1	-1	0	1	-1	0	1	-1	0	1	
S1	-3	-3	-3	1	1	1	1	1	1	1	1	1	
S2	0	0	0	-2	-2	-2	1	1	1	1	1	1	
S3	0	0	0	0	0	0	-1	-1	-1	1	1	1	
													L
L×S1	+3	0	-3	-1	0	1	-1	0	1	-1	0	1	-12.0
L×S2	0	0	0	+2	0	-2	-1	0	1	-1	0	1	- 9.0
L×S3	0	0	0	0	0	0	+1	0	-1	-1	0	1	- 3.0

$SS_{Lin \times Subj} = SS_{L\times S1} + SS_{L\times S2} + SS_{L\times S3} = (1\times-12^2)/24 + (1\times-9^2)/12 + (1\times-3^2)/4 = 6.0+6.75+2.25 = 15.0$

$df_{Lin\times Subj} = 1 + 1 + 1 = 3$ $MS_{ErrLin} = 15.0 / 3 = 5.0$

CHAPTER 8 - WITHIN-S FACTORIAL

In a Within-S factorial design with A levels to factor A and B levels to factor B, observations in all A×B cells are expected to correlate. In a cognitive study of two factors that influence detection of spelling errors, for example, six participants read six passages containing 20 spelling errors. Three passages contained easy errors (e.g., ryte for write), and three passages difficult errors (e.g., right for write). For each difficulty level, one of three passages was read without interference, one read with low interference (soft noise in the background), and one read with high interference (loud speech in the background). Detection of spelling errors was expected to be higher for easy mistakes and to decrease with increasing interference, especially for difficult errors.

		Easy			Difficult		
		None	Low	High	None	Low	High
Subject	1.	9	9	8	12	6	3
	2.	10	10	9	12	8	4
	3.	9	9	8	9	6	2
	4.	8	6	7	9	6	3
	5.	7	7	6	8	5	1
	6.	8	7	7	7	5	2

The Difficulty (Easy, Difficult) and Interference (None, Low, High) factors are both Within-S because each person contributed scores to all 2×3 = 6 cells or conditions. Scores are expected to correlate across the six conditions given participants who recognize many or few errors in one condition should perform similarly across conditions relative to other subjects. This study would also be a Within-S factorial if 36 people were matched on a pretest of spelling ability, sorted into groups of six people with similar pretest scores, and one person from each grouping of similar-ability people was randomly assigned to each condition. Under these circumstances, scores are again expected to correlate across conditions, which warrants a Within-S factorial analysis. Given 6 subjects and 6 scores per subject, there are N=36 observations.

A Within-S analysis considers Subjects (S) to be a third factor in the study. Given three factors, there are three main effects (D, I, S), three two-way interactions (D×I, D×S, I×S), and one three-way interaction (D×I ×S). That is, SS_{Total} is partitioned as follows:

$$SS_{Total} = SS_A + SS_B + SS_{A\times B} + SS_S + SS_{A\times S} + SS_{B\times S} + SS_{A\times B\times S}$$

Basic Calculations for Within-S Factorial

The three numerators for this design are calculated the same as for a Between-S or a Mixed factorial (i.e., one Within-S and one Between-S factor discussed later). Cell means are arranged as previously and calculations illustrated below.

	$\bar{y}_{di}$ n_{di} = 6	Interference			n_d = 18	
		None	Low	High	$\bar{y}_d$	$\bar{y}_d$ $\bar{y}_G$
Difficulty	Easy	8.5	8.0	7.5	8.0	+1.0
	Difficult	9.5	6.0	2.5	6.0	−1.0
n_i=12	$\bar{y}_i$	9.0	7.0	5.0	$\bar{y}_G$ = 7.0	
	$\bar{y}_i - \bar{y}_G$	+2.0	0.0	−2.0	N = 36	

SS_{Diff} = 36.0 = 18(+1² + −1²) df = 2−1 = 1

SS_{Int} = 96.0 = 12(+2² + 0² + −2² df = 3−1 = 2

$\bar{y}_G + (\bar{y}_d - \bar{y}_G) + (\bar{y}_i - \bar{y}_G) = \bar{y}_{di}'$

		N	L	H
Diff	E	10	8	6
	D	8	6	4

$\bar{y}_{di} - \bar{y}_{di}'$

	N	L	H
E	−1.5	0	+1.5
D	+1.5	0	−1.5

$SS_{D\times I}$ = 54.0 = 6×(−1.5² + 0² + 1.5² + 1.5² + 0² + −1.5²) df=(2−1)×(3−1) = 2

MS_D = 36.0/1 = 36.0 MS_I = 96.0/2 = 48.0 MS_{DxI} = 54.0/2 = 27.0

The interaction effects above are deviations of observed cell means from cell means predicted from just main effects. Alternatively, $SS_{D\times I}$ could be obtained by subtracting main effects from cell means to produce deviations from $\bar{y}_G$. SSs for the two main effects and the interaction are divided by their df to produce MSs that are numerators for three omnibus tests: main effect of Difficulty, main effect of Interference, and the Difficulty by Interference interaction.

Instead of a single denominator for these tests as for Between-S factorials, Within-S factorials first remove variability due to the main effect of Subjects (SS_S) and then calculate separate interactions with Subjects as appropriate error terms: $SS_{D\times S}$ is the denominator for the Difficulty main effect, $SS_{I\times S}$ for the Interference main effect, and $SS_{D\times I\times S}$ for the Difficulty by Interference interaction. We first calculate SS_S.

		Easy			Difficult			n_s = 6	
		None	Low	High	None	Low	High	$\bar{y}_s$	$\bar{y}_s - \bar{y}_G$
Subj	1.	9	9	8	12	6	3	7.8333	.833
	2.	10	10	9	12	8	4	8.8333	1.833
	3.	9	9	8	9	6	2	7.1667	.167
	4.	8	6	7	9	6	3	6.5000	-.500
	5.	7	7	6	8	5	1	5.6667	-1.333
	6.	8	7	7	7	5	2	6.0000	-1.000

SS_S = 6×(.833² + 1.833² + ... + -1.00²) = 42.67 df = 6 - 1 = 5

Calculating the three interaction terms involving Subjects is labour intensive, especially for studies with many subjects. Even with just six subjects here, the Difficulty by Subject interaction requires 2×6 = 12 cell means averaged over the three levels of the Interference factor. This generates an Easy and Difficult mean for each subject based on three scores ($\bar{y}_{ds}$, n_{ds} = 3). For subject one, the Easy mean would be (9+9+8)/3 = 5.967, and the Difficult mean (12+6+3)/3 = 7.00. Formulas for the interaction could be used to create deviation scores for each of the 12 cells, which would be squared, summed, and multiplied by 3 to get $SS_{D\times S}$, df = (2 - 1)(6 - 1) = 5.

The Interference by Subject interaction requires 3×6 = 18 cell means averaged across the two levels of Difficulty. This would generate a None, Low, and High mean for each subject based on two scores ($\overline{y}_{is}$, n_{is} = 2). For subject one, the None mean would be (9+12)/2 = 10.50, the Low mean would be (9+6)/2 = 7.50, and the High mean would be (8+3)/2 = 5.50. Interaction deviations would be calculated for each of the 12 cells, squared, summed, and multiplied by 2 to get $SS_{I×S}$, df = (3 - 1)(6 - 1) = 10.

The three way Difficulty by Interference by Subject interaction would most easily be calculated by subtracting all the preceding sources of variability from SS_{Total}, as shown here. Happily SPSS procedures determine the appropriate error terms, avoiding the preceding complications. It is important to appreciate, however, what the denominators for Within-S effects represent.

$SS_{D×I×S} = SS_{Total} - SS_D - SS_I - SS_{D×I} - SS_S - SS_{D×S} - SS_{I×S}$ df = (2-1)(3-1)(6-1) = 10

SPSS Omnibus Analyses for Within-S Factorial

Data for Within-S Factorial designs is generally entered as multiple scores (variables) for each subject, with one score for each cell in the A×B design. Here six scores are entered per person. Although names for variables are arbitrary, it helps to label scores to indicate what levels of the two factors they correspond to. This helps users specify for MANOVA and GLM which cell each score belongs to.

```
DATA LIST FREE / easnone easlow eashigh difnone diflow difhigh.
BEGIN DATA
 9  9  8    12  6  3          10 10  9    12  8  4           9  9  8     9  6  2
 8  6  7     9  6  3           7  7  6     8  5  1           8  7  7     7  5  2
END DATA.
```

How to assign scores to cells for SPSS is illustrated below. The variables are listed after the MANOVA or GLM command in a certain order; the order below corresponds to their order in the dataset. Scores are listed (shown below or in the order of the descriptive statistics produced by MANOVA) so that the Difficulty factor changes from one level to the next more slowly than the Interference factor. That is, Interference is said to be "nested" within levels of Difficulty. A /WSF option specifies how scores relate to the six cells. Here, the /WSF option specifies Difficulty first (*diff* with two levels) followed by Interference (*int* with three levels). If scores were listed with Difficulty nested within Interference, then the listing of factors in the /WSF option would have to correspond to the new order. To illustrate, the correct SPSS command would be: MANOVA easnone difnone easlow diflow eashigh difhigh /WSF = int(3) diff(2). SSs and dfs below agree with earlier calculations.

```
MANOVA easnone easlow eashigh difnone diflow difhigh
   /WSF = diff(2) int(3) /PRINT = CELL.
```

	Mean	Std. Dev.	*Diff*	*Int*	
Variable .. easnone	8.500	1.049	*1*	*1*	
Variable .. easlow	8.000	1.549	*1*	*2*	
Variable .. eashigh	7.500	1.049	*1*	*3*	***Int nested in Diff***
Variable .. difnone	9.500	2.074	*2*	*1*	
Variable .. diflow	6.000	1.095	*2*	*2*	
Variable .. difhigh	2.500	1.049	*2*	*3*	

Tests of Between-Subjects Effects.

Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	42.67	5	8.53			$= SS_S$
CONSTANT	1764.00	1	1764.00	206.72	.000	

Tests involving 'DIFF' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	4.00	5	.80			$= SS_{D\times S}$
DIFF	36.00	1	36.00	45.00	.001	$= SS_D$

Tests involving 'INT' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	4.33	10	.43			$= SS_{I\times S}$
INT	96.00	2	48.00	110.77	.000	$= SS_I$

Tests involving 'DIFF BY INT' Within-Subject Effect.

Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	5.00	10	.50			$SS_{D\times I\times S}$
DIFF BY INT	54.00	2	27.00	54.00	.000	$S_{D\times I}$

The syntax for GLM is similar to that for MANOVA, as shown later. With menus, select *Analyze* | *GLM* | *Repeated Measures* to bring up the screen in Figure 8-1. Specify the two Within-S factors and the number of levels for each. In Figure 8-1, *diff* has already been added to the design and *int* is ready to be added.

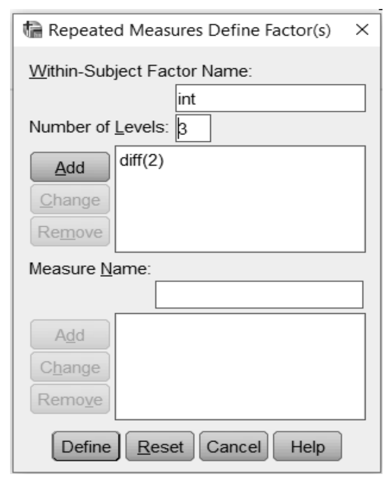


Figure 8-1. Within-S Menu.

Select *Define* to proceed to the next screen shown in Figure 8-2. An empty frame has been created for both Within-S factors and scores must be moved into the proper cells of the design. Whether these can be entered individually or as a set depends on the order of the variables in the dataset and the order of the cells in the design. In the present example, the order of variables in the data and the frame correspond, so they all could be selected and moved together into the slots. If *int* and *diff* were created in the opposite order on the preceding screen, it would be necessary to move variables one at a time.

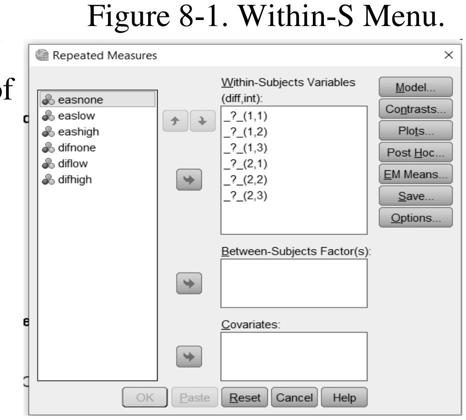


Figure 8-2. Within-S Step 2

After the six variables have been placed in the frame and any additional aspects of the design specified, Click *OK* to run the analysis.

Given default values are included, the syntax generated will be more complex than required, as shown below.

```
GLM easnone easlow eashigh difnone diflow difhigh
  /WSFACTOR = diff 2 Polynomial  int  3 Polynomial
  /METHOD = SSTYPE(3)  /CRITERIA = ALPHA(.05)
  /WSDESIGN = diff int diff*int .
```

By default for Within-S designs, polynomial contrasts have been requested for both *diff* and *int*, even though the former has only two levels. The default /WSDESIGN is a full factorial, with main effects for *diff* and *int* and the *diff**int* interaction (*diff**int* is equivalent to *diff* BY *int*). GLM output is shown later.

Planned Contrasts for Main Effects

As for the Between-S factorial, follow-up analyses are generally required to interpret main and interaction effects with df > 1. Both main effects are significant here but Difficulty does not require additional analyses given *df* = 1. Given *df* = 2, the Interference main effect and the D×I interaction benefit from further analyses. The various follow-up analyses involve calculations for the numerators that are identical to the Between-S design, but different error terms are required. We start with planned contrasts.

Linear and quadratic contrasts are calculated below for the Interference effect, first using main effect means averaged across Difficulty, and then using cell means. The correct n is important for the calculations and is the number of observations per mean, not necessarily number of subjects. The main effect means for Interference, for example, are based on two observations for each of six subjects, leading to $n_i = 12$, whereas the cell means involve one observation per subject, hence, $n_{di} = 6$.

			Interference			
		None	Low	High		
Diff	Easy	8.5	8.0	7.5		
	Difficult	9.5	6.0	2.5		
	$\overline{y}_i$	*9.0*	*7.0*	*5.0*	*L*	*SS* $n_s = 12$
	Lin	-1	0	1	-4.0	96.0 = 12×-4²/2
	Qua	-1	2	-1	0.0	0.0
					Sum =	96.0 = $SS_{Interference}$

OR

	Easy			Difficult				
	None	Low	High	None	Low	High		
$\overline{y}_{di}$	*8.5*	*8.0*	*7.5*	*9.5*	*6.0*	*2.5*	*L*	*SS* $n_s = 6$
Lin	-1	0	1	-1	0	1	-8.0	96.0 = 6×-8²/4
Qua	-1	2	-1	-1	2	-1	0.0	0.0 = 6×0²/12

The following MANOVA commands are equivalent to those for the single factor Within-S design. The linear effect of Interference accounts for all of the variability due to interference and the quadratic contrast accounts for none because the scores decrease in a strictly linear manner from None to Low to High. That is, the means of 9.0, 7.0, and 5.0 decrease by two units for every one unit increase in Interference. The analysis also partitions the D×I interaction, discussed shortly.

Note below that the analysis has also partitioned the I×S denominator for the Interference effect resulting in different denominators for the linear and quadratic effects. That is, $SS_{I\times S} = 4.33 = 2.500 + 1.833$.

```
MANOVA easnone TO difhigh /WSF = diff(2) int(3)
   /PRINT = SIGN(UNIV)  /CONTR(int) = POLY.
...
Tests involving 'INT' Within-Subject Effect.
 Univariate F-tests with (1,5) D. F.
 Variable  Hypoth. SS    Error SS  Hypoth. MS   Error MS          F    Sig. of F
 T3          96.00000     2.50000    96.00000     .50000  192.00000         .000
 T4            .00000     1.83333      .00000     .36667     .00000        1.000

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS                4.33     10       .43                        SS_I×S
 INT                        96.00      2     48.00    110.77      .000

 Estimates for T3 --- Individual univariate .9500 confidence intervals
  Parameter           Coeff.   Std. Err.       t-Value           Sig. t
       1      -4.0000000000      .28868     -13.85641           .00004

 Estimates for T4 --- Individual univariate .9500 confidence intervals
  Parameter           Coeff.   Std. Err.       t-Value           Sig. t
       1        .0000000000      .24721        .00000          1.00000

Tests involving 'DIFF BY INT' Within-Subject Effect.
 Univariate F-tests with (1,5) D. F.
 Variable     Hypoth. SS    Error SS    Hypoth. MS   Error MS           F    Sig. of F
 T5             54.00000    2.50000       54.00000     .50000   108.00000         .000
 T6               .00000    2.50000         .00000     .50000      .00000        1.000

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS                5.00     10       .50
 DIFF BY INT                54.00      2     27.00     54.00      .000
```

Within-S contrasts were conceptualized earlier as individual contrast scores for each Subject, with a test of the significance of the difference between the mean contrast score and 0 relative to the variability in the contrast scores. This is illustrated below, using normalized coefficients to increase correspondences with the analysis of variance. Integer contrast coefficients would produce the same final *F*s and *p*s, but different intermediate values (e.g., SSs). Note below that $SS_{Linear} = 6\times(-4 - 0)^2 = 96.0$ and $SS_{L\times S} = (6 - 1)\times.707^2 = 2.50$. As before, the test is a generalization of the paired difference t -test.

```
COMPUTE intlin = -.5*easnone+0*easlow+.5*eashigh+-.5*difnone+0*diflow+.5*difhigh.
LIST.
easnone easlow eashigh difnone diflow difhigh         intlin
   9.00   9.00    8.00   12.00   6.00    3.00         -5.00
  10.00  10.00    9.00   12.00   8.00    4.00         -4.50
   9.00   9.00    8.00    9.00   6.00    2.00         -4.00
   8.00   6.00    7.00    9.00   6.00    3.00         -3.50
   7.00   7.00    6.00    8.00   5.00    1.00         -4.00
   8.00   7.00    7.00    7.00   5.00    2.00         -3.00
```

```
MANOVA intlin /PRINT = CELL.
                                                           Mean  Std. Dev.    N
 For entire sample                                       -4.000       .707    6

 Source of Variation           SS      DF         MS         F  Sig of F
 WITHIN CELLS                2.50       5        .50
 CONSTANT                   96.00       1      96.00    192.00      .000
                        SSL=6×(-4.0-0)²                 t = 4/(.707/√6) = 13.859 = √F
```

Planned contrasts are generated automatically for GLM, with the default contrasts being polynomial. The output is shown below. Different contrasts would be specified following the factor name in the /WSF option. As for contrasts in the single factor Within-S design, SSs for the partitioned effect may depend on whether normalized or integer coefficients are used. For *SPECIAL* contrasts, k 1s must precede the k - 1 sets of k coefficients that represent the contrast.

```
GLM easnone TO difhigh /WSF = diff(2) int(3).
...
Tests of Within-Subjects Contrasts
 Source           diff   int       Type III Sum of df Mean Square F        Sig.
                                   Squares
 diff             Linear           36.000          1  36.000      45.000  .001
 Error(diff)      Linear            4.000          5    .800

 int                     Linear    96.000          1  96.000     192.000   .000
                         Quadratic   .000          1    .000        .000  1.000
 Error(int)              Linear     2.500          5    .500
                         Quadratic  1.833          5    .367

 diff * int       Linear Linear    54.000          1  54.000     108.000   .000
                         Quadratic   .000          1    .000        .000  1.000
 Error(diff*int) Linear Linear      2.500          5    .500
                         Quadratic  2.500          5    .500
...
```

Post Hoc Comparisons for Main Effects

Although GLM does not provide a full range of post hoc tests for Within-S factors, /EMMEANS can be used to obtain LSD or BONFERRONI tests, as illustrated below for the default LSD option.

```
GLM easnone TO difhigh /WSF = diff(2) int(3)
  /EMMEANS = TABLES(int) COMPARE(int) ADJ(LSD).
...
Estimated Marginal Means
 int  Mean  Std.  95% Confidence Interval
            Error Lower Bound      Upper Bound
 1    9.000 .606  7.443            10.557
 2    7.000 .500  5.715            8.285
 3    5.000 .408  3.951            6.049

Pairwise Comparisons
 (I)  (J)  Mean Difference Std.  Sig.(a)
 int  int  (I-J)           Error

 1    2    2.000(*)        .258  .001      t=2.0/.258 = 7.752
      3    4.000(*)        .289  .000
 2    3    2.000(*)        .258  .001
```

In the single factor Within-S design, the post hoc tests were essentially paired difference t-tests. This is also true for the factorial design, but using Interference means averaged across the levels of Difficulty. Except for rounding, t values correspond to ts obtained by dividing *Mean Difference* by *Std. Error* in the preceding GLM, and the p values in the analyses correspond. Although frowned on by some statisticians, *t*s could be converted to *q*s to conduct SNK or Tukey tests.

```
COMPUTE none = MEAN(easnone, difnone).
COMPUTE low  = MEAN(easlow, diflow).
COMPUTE high = MEAN(eashigh, difhigh).

TTEST PAIRED none low high.
...
```

	Paired Differences			t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean			
Pair 1 none - low	2.00000	.63246	.25820	**7.746**	**5**	**.001**
Pair 2 none - high	4.00000	.70711	.28868	13.856	5	.000
Pair 3 low - high	2.00000	.63246	.25820	7.746	5	.001

Follow-Up Analyses for the Interaction

Follow-up analyses for the interaction include partitioning the interaction, simple effects, or both. Calculations of numerators are the same as for Between-S factorial designs, but unique error terms are required. As with previous Within-S contrasts, the unique error terms are best thought of as variability in contrast scores for each subject. Partitioning the interaction is shown first. Partitioning of the numerator is illustrated below.

		Easy			Difficult				
		None	Low	High	None	Low	High		
$\bar{y}_{di}$		8.5	8.0	7.5	9.5	6.0	2.5	L	SS
Main Effect of Difficulty									
	Dif	−1	−1	−1	1	1	1		
Main Effect of Interference									
	Lin	−1	0	1	−1	0	1		
	Qua	−1	2	−1	−1	2	−1		
Difficulty × Interference Interaction									
	D×Lin	1	0	−1	−1	0	1	−6.0	$54.0 = 6\times6^2/4$
	D×Qua	1	−2	1	−1	2	−1	0.0	0.0
									$\sum = 54.0 = SS_{D\times I}$

The partitioning of the interaction was produced automatically in earlier MANOVA and GLM analyses. Relevant output is presented below. The denominators and their *df* have been partitioned as well as the numerators. These unique denominators, which happen to be equal in this example, represent variability in individual contrast scores, as shown following the MANOVA.

```
MANOVA easnone TO difhigh /WSF=diff(2) int(3) /PRINT=SIGN(UNIV) /CONTR(int)=POLY.
...
Tests involving 'DIFF BY INT' Within-Subject Effect.
 Univariate F-tests with (1,5) D. F.
 Variable    Hypoth. SS    Error SS  Hypoth. MS    Error MS            F    Sig. of F
 T5            54.00000    2.50000   54.00000       .50000    108.00000         .000
 T6              .00000    2.50000     .00000       .50000       .00000        1.000
...

*Normalized coefficients.
COMPUTE difxlin = .5*easnone+0*easlow+-.5*eashigh+-.5*difnone+0*diflow+.5*difhigh.

LIST easnone TO difhigh difxlin.

easnone easlow eashigh difnone diflow difhigh difxlin
   9.00   9.00    8.00   12.00   6.00    3.00   -4.00
  10.00  10.00    9.00   12.00   8.00    4.00   -3.50
   9.00   9.00    8.00    9.00   6.00    2.00   -3.00
   8.00   6.00    7.00    9.00   6.00    3.00   -2.50
   7.00   7.00    6.00    8.00   5.00    1.00   -3.00
   8.00   7.00    7.00    7.00   5.00    2.00   -2.00

MANOVA difxlin /PRINT = CELL.

 Variable .. difxlin                             Mean  Std. Dev.    N
 For entire sample                             -3.000       .707    6

 Source of Variation          SS       DF         MS          F  Sig of F
 WITHIN CELLS               2.50        5        .50                       SSDxLxS = .707²(6-1)
 CONSTANT                  54.00        1      54.00     108.00      .000  SSDxL = 6(-3.0-0)²

GLM easnone TO difhigh /WSF = diff(2) int(3).
...
Tests of Within-Subjects Contrasts
...
 Source           diff   int       Type III Sum of df Mean Square    F       Sig.
                                   Squares
 diff * int       Linear Linear    54.000          1  54.000      108.000   .000
                         Quadratic   .000          1    .000         .000  1.000
 Error(diff*int)  Linear Linear     2.500          5    .500
                         Quadratic  2.500          5    .500
...
```

A second way to partition the interaction in GLM uses /MMATRIX. List the variables corresponding to the six cells, with each score followed by the appropriate contrast coefficient, normalized here so that SSs agree with earlier analyses as well as the final tests of significance.

```
GLM easnone TO difhigh /WSF = diff(2) int(3)
  /MMATRIX easnone .5 easlow 0 eashigh -.5 difnone -.5 diflow 0 difhigh .5.
...

 L1          Contrast Estimate            -3.000
             Std. Error                    .289                t = 3.0/.289 = 10.381
             Sig.                          .000

 Source   Sum of Squares df Mean Square F        Sig.
 Contrast 54.000         1  54.000      108.000 .000           F = 10.381² = t²
 Error    2.500          5  .500
```

To appreciate why the Difficulty by Linear component of the interaction accounts entirely for $SS_{D\times I}$, compare the interaction deviations used to calculate $SS_{D\times I}$ to the contrast coefficients (or their reverse). The interaction deviations below correspond perfectly with contrast coefficients (the integers below with signs changed to make comparison easier).

		Interference					
		None	Low	High		$\bar{y}_d$	$\bar{y}_d - \bar{y}_G$
Difficulty	Easy	8.5	8.0	7.5		*8.0*	*+1.0*
	Difficult	9.5	6.0	2.5		*6.0*	*-1.0*
	$\bar{y}_i$	*9.0*	*7.0*	*5.0*	$\bar{y}_G$	*7.0*	
	$\bar{y}_i - \bar{y}_G$	*+2.0*	*0*	*-2.0*			

Expected from Main Effects			Difference = Interaction			Interaction Contrast Coeff.		
10.0	8.0	6.0	-1.5	0	+1.5	-1	0	1
8.0	6.0	4.0	+1.5	0	-1.5	1	0	-1

Simple Effects Analysis for Within-S Interaction

The second approach to follow-up for interactions tests the simple effects of one factor at each level of the other factor. The calculations below determine the overall simple effect of *int* at the Difficult level of *diff*, and partition the *df* = 2 simple effect into linear and quadratic components.

	Interference			
	None	Low	High	$\bar{y}_d$
Difficult	9.5	6.0	2.5	*6.0*
$\bar{y}_{di} - \bar{y}_d$	-3.5	0	+3.5	

$SS_{IwDifficult} = 6\times(-3.5^2 + 0^2 + 3.5^2) = 147.0$

	Easy			Difficult				
	None	Low	High	None	Low	High		
$\bar{y}_{di}$	8.5	8.0	7.5	9.5	6.0	2.5	L	SS
Lin	0	0	0	-1	0	1	-7.0	$147.0 = 6\times -7^2/2$
Qua	0	0	0	-1	2	-1	0.0	0.0
								Sum = 147.0 = $SS_{IwDifficult}$

The following MANOVA produces the overall simple effects analysis for Easy and Difficult conditions, as well as numerators and denominators partitioned into linear and quadratic components of the simple effects. One anomaly discussed later is the lack of error variability for the linear effect of Interference for Easy items. As for Between-S factorial designs, simple effects analyses represent a new partitioning of SS_{Total}, specifically: $SS_{IwD(1)} + SS_{IwD(2)} = 3.0 + 147.0 = 150.0 = SS_I + SS_{D\times I} = 96.0 + 54.0$.

```
MANOVA easnone TO difhigh /WSF = diff(2) int(3) /PRINT = SIGN(UNIV)
  /CONTR(diff) = POLY  /WSD = diff int WITHIN diff(1) int WITHIN diff(2).
```

```
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           42.67       5      8.53
 CONSTANT                1764.00       1   1764.00    206.72      .000

Tests involving 'DIFF' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL            4.00       5       .80
 DIFF                      36.00       1     36.00     45.00      .001
```

```
Tests involving 'INT WITHIN DIFF(1)' Within-Subject Effect.
 Univariate F-tests with (1,5) D. F.
Variable    Hypoth. SS    Error SS      Hypoth. MS    Error MS                  F       Sig. of F
T3             3.00000      .00000         3.00000      .00000   3377699720527874            .000
T4              .00000     2.33333          .00000      .46667             .00000           1.000

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL            2.33      10       .23
 INT WITHIN DIFF(1)         3.00       2      1.50      6.43      .016
```

```
Tests involving 'INT WITHIN DIFF(2)' Within-Subject Effect.
 Univariate F-tests with (1,5) D. F.
 Variable     Hypoth. SS     Error SS     Hypoth. MS     Error MS           F    Sig. of F
 T5            147.00000      5.00000      147.00000      1.00000   147.00000         .000
 T6               .00000      2.00000         .00000       .40000      .00000        1.000

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL            7.00      10       .70
 INT WITHIN DIFF(2)       147.00       2     73.50    105.00      .000
```

Linear contrast scores for the simple effects reveal the nature of the error terms and also why the linear effect of Interference for Easy items showed no variability. Here are the relevant computations. Everybody has exactly the same score for *linweasy*; there is no variability in these scores. This occurs because *eashigh* is one unit lower than *easnone* for all six participants.

```
COMPUTE linwdiff = 0*easnone+0*easlow+0*eashigh+-.7071*difnone+0*diflow+.7071*difhigh.

COMPUTE linweasy = -.7071*easnone+0*easlow+.7071*eashigh+0*difnone+0*diflow+0*difhigh.

LIST easnone TO difhigh linwdiff linweasy.
easnone easlow eashigh difnone diflow difhigh linwdiff linweasy
 9.0000 9.0000  8.0000  12.000 6.0000  3.0000  -6.364   -.7071
 10.000 10.000  9.0000  12.000 8.0000  4.0000  -5.657   -.7071
 9.0000 9.0000  8.0000  9.0000 6.0000  2.0000  -4.950   -.7071
 8.0000 6.0000  7.0000  9.0000 6.0000  3.0000  -4.243   -.7071
 7.0000 7.0000  6.0000  8.0000 5.0000  1.0000  -4.950   -.7071
 8.0000 7.0000  7.0000  7.0000 5.0000  2.0000  -3.536   -.7071

MANOVA linwdiff linweasy /PRINT = CELL.
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 *   W A R N I N G   * These variables have NO variance ...              *
 *                   *   linweasy                                        *
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

```
                           Mean  Std. Dev.     N
 Variable .. linwdiff   -4.950      1.000     6
 Variable .. linweasy    -.707       .000     6
```

SE=1.0/√6=.408   ***t=4.95/.408=12.13***

$t^2=12.13^2=147.14\approx F$

```
 Univariate F-tests with (1,5) D. F.
 Variable    Hypoth. SS    Error SS   Hypoth. MS    Error MS                  F    Sig. of F
 linwdiff     146.99718     4.99990    146.99718      .99998          147.00000         .000
 linweasy       2.99994      .00000      2.99994      .00000  1.415024692E+31         .000
```

The standard F test for the overall simple effects of Within-S factors cannot be obtained by /EMMEANS, as done for the Between-S factorial. GLM instead provides Multivariate statistics for the simple effects of Within-S factors, along with pairwise comparisons at each level of the other factor.

```
GLM  easnone easlow eashigh difnone diflow difhigh
  /WSFACTOR = diff 2 Polynomial int 3 Polynomial
  /EMMEANS = TABLES(diff*int) COMPARE(int).
...
Estimated Marginal Means
diff * int
Pairwise Comparisons
 diff (I)   (J)  Mean Difference Std.  Sig.(a)
      int   int  (I-J)           Error

 1    1     2    .500            .342  .203
            3    1.000           .000  .
      2     3    .500            .342  .203

 2    1     2    3.500(*)        .563  .002
            3    7.000(*)        .577  .000
      2     3    3.500(*)        .224  .000

Multivariate Tests
 diff                   Value  F          Hypothesis df Error df Sig.
 1    Pillai's trace    .300   2.143(a)   1.000         5.000    .203
      Wilks' lambda     .700   2.143(a)   1.000         5.000    .203
      Hotelling's       .429   2.143(a)   1.000         5.000    .203
      trace
      Roy's largest     .429   2.143(a)   1.000         5.000    .203
      root
...
```

One way to do a simple effects analysis in GLM is to conduct two single-factor analyses, one for each level of Difficulty. This analysis is shown below for the Difficult level of the Difficulty factor. Note the correspondence to the MANOVA results.

```
GLM  difnone diflow difhigh /WSFACTOR = int 3 Polynomial.
...
Tests of Within-Subjects Effects
 Source                      Type III Sum of df    Mean Square F       Sig.
                             Squares
 int         Sphericity      147.000         2     73.500      105.000 .000
 Error(int)  Sphericity      7.000           10    .700
```

```
Tests of Within-Subjects Contrasts
 Source      int       Type III Sum of df Mean Square F       Sig.
                       Squares
 int         Linear    147.000         1  147.000     147.000 .000
             Quadratic .000            1  .000        .000    1.000
 Error(int)  Linear    5.000           5  1.000
             Quadratic 2.000           5  .400

Tests of Between-Subjects Effects
 Source    Type III Sum of df Mean Square F       Sig.
           Squares
 Intercept 648.000         1  648.000     124.615 .000
 Error     26.000          5  5.200
```

The /MMATRIX option in GLM can also be used to obtain simple effects, as shown below.

```
GLM easnone easlow eashigh difnone diflow difhigh /WSF = diff(2) int(3)
  /MMATRIX easnone 0 easlow 0 eashigh 0 difnone -.7071 diflow 0 difhigh .7071.
...
Custom Hypothesis Tests
 L1          Contrast Estimate          -4.950         t = 4.95/.408 = 12.13
             Std. Error                   .408
             Sig.                         .000

 Source   Sum of Squares df Mean Square F       Sig.
 Contrast 146.997        1  146.997     147.000 .000    √F = 12.12 = t
 Error    5.000          5  1.000
```

Conclusions

The Within-S factorial presents a few complications, notably in determining appropriate error terms. All operations on the numerator side are calculated exactly as for Between-S factorials, including main effects and interaction for the omnibus ANOVA, follow-up analyses of main effects, and both partitioning and simple effects approaches to the interaction. This is also the case for mixed designs, involving one Within-S and one Between-S factor. Given the appropriate commands for the design, SPSS computes the proper error terms.

Appendix 8-1: Within-S Factorial in Between-S Format

Below the data is entered and re-analyzed in Between-S format; that is, the data file contains one observation per case along with codes for Difficulty (*diff*), Interference (*int*), and Subject (*subj*).

```
*Data in Between-s format.
DATA LIST FREE / subj diff int score.
BEGIN DATA
  1 1 1  9    1 1 2  9    1 1 3  8    1 2 1 12    1 2 2  6    1 2 3  3
  2 1 1 10    2 1 2 10    2 1 3  9    2 2 1 12    2 2 2  8    2 2 3  4
  3 1 1  9    3 1 2  9    3 1 3  8    3 2 1  9    3 2 2  6    3 2 3  2
  4 1 1  8    4 1 2  6    4 1 3  7    4 2 1  9    4 2 2  6    4 2 3  3
  5 1 1  7    5 1 2  7    5 1 3  6    5 2 1  8    5 2 2  5    5 2 3  1
  6 1 1  8    6 1 2  7    6 1 3  7    6 2 1  7    6 2 2  5    6 2 3  2
END DATA.
```

The MANOVA below requests the full factorial, which uses all degrees of freedom. It shows seven terms from the Within-S factorial, including the Subj by Diff by Int three-way interaction. Find the rows that match the previous MANOVA and later GLM output.

```
MANOVA score BY subj(1 6) diff(1 2) int(1 3).
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 *   W A R N I N G   *  Too few degrees of freedom in RESIDUAL          *
 *                   *  error term for the following test(s) (DF = 0). *
 * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
 Source of Variation          SS      DF        MS         F  Sig of F
 RESIDUAL                    .00       0        .
 subj                      42.67       5      8.53        .         .
 diff                      36.00       1     36.00        .         .
 int                       96.00       2     48.00        .         .
 subj BY diff               4.00       5       .80        .         .
 subj BY int                4.33      10       .43        .         .
 diff BY int               54.00       2     27.00        .         .
 subj BY diff BY int        5.00      10       .50        .         .

 (Model)                  242.00      35      6.91        .         .
 (Total)                  242.00      35      6.91
```

With data in Between-S format, GLM can calculate the three-way Difficulty by Interference by Subjects interaction. $SS_{D\times I\times S}$ equals the squared deviations of observed scores from values predicted from three main effects and three two-way interactions, as shown below.

```
GLM score BY diff int subj
 /DESIGN diff int subj   diff BY int   diff BY subj   int BY subj
 /SAVE PRED(prd2way).
...
COMPUTE DxIxS = score - prd2way.
COMPUTE DxIxS2 = dis**2.

DESCR DxIxS2 /STAT = SUM.
                 N  Sum
 dis2           36 5.0000          SSDxIxS
```

Chapter 9 - Mixed Factorial ANOVA

A mixed-factorial design (also known as a split-plot design) involves at least two factors, one or more Within-S factors and one or more Between-S factors. The proper analysis combines Between-S and Within-S designs. In the present study, educational psychologists examined the effects on later reading performance of a pre-school program (*prog*): program one was a control group with no pre-school intervention, program two was a pre-school group that emphasized social experiences rather than academic preparedness, and program three was an academic preparedness pre-school program that developed skills relevant to later reading (e.g., phonics awareness). Fifteen schools from disadvantaged areas were randomly assigned to conditions (5 schools per program). Reading tests were administered prior to the intervention (Pre or Time 1), in Grade 1 (G1 or Time 2), and in Grade 5 (G5 or Time 3). Researchers predicted that only Prog 3 would affect reading performance and that the benefits would persist across grades. In this study, the five schools in each condition are "subjects," program is Between-S, and time of testing is Within-S.

Calculations for Default Mixed Factorial ANOVA

Numerator SSs and dfs for main effects and interaction are calculated as for all factorial designs. The means for the nine cells (3×3) permit calculation of the numerators.

	$\bar{y}_{pt}$	$n_{pt} = 5$		Time			
		Pre	G1	G5		$\bar{y}_p$	$\bar{y}_p - \bar{y}_G$ $n_p = 15$
	1	11.4	10.0	11.6		*11.000*	*.311*
Prog	2	9.0	10.2	7.6		*8.933*	*-1.756*
	3	9.6	13.6	13.2		*12.133*	*1.444*
	$\bar{y}_t$	*10.000*	*11.267*	*10.800*	$\bar{y}_G = 10.689$		
	$\bar{y}_t - \bar{y}_G$	*-.689*	*.578*	*.111*			
	$n_t = 15$						

$SS_{Prog} = 78.98 = 15(.311^2 + -1.756^2 + 1.444^2)$ $\quad df = 3 - 1 = 2$

$SS_{Time} = 12.32 = 15(-.689^2 + .578^2 + .111^2)$ $\quad df = 3 - 1 = 2$

$SS_{P\times T} = 60.76 = 5[\{11.4 - (10.689 + .311 - .689)\}^2 + \{10.0 - (10.689 + .311 + .578)\}^2 \ldots]$

$= 5[\{(11.4 - -.689 - .311) - 10.689\}^2 + \{10.0 - .578 - -1.756\} - 10.689\}^2 \ldots]$

$= 5(1.089^2 + .756^2 + -1.844^2 + -1.578^2 + .689^2 + .889^2 + .489^2 + -1.444^2 + .956^2)$

$df = (3 - 1)(3 - 1) = 4$

A Between-S Error (Subjects Within Program) term is used for the Between-S factor *prog* and a Within-S Error (Time By Subjects Within Program) for the Within-S factor *time* and the interaction. $SS_{Subjects}$ cannot be calculated across all 15 subjects (schools) because the five schools in each program are unrelated and $SS_{Subjects}$ would include program variability. It is appropriate, however, to calculate variability due to Subjects *Within* each Program level (i.e., SS_{SwP1}, SS_{SwP2}, SS_{SwP3}) and aggregate these SSs to obtain SS_{SwP}, as done for the Between-S single factor design. SS_{SwP} has df = (5-1) + (5-1) + (5-1) = 12 and provides the error for *prog*.

It is also inappropriate to calculate an overall Time by Subject interaction, because variability associated with both Program and the Time by Program interaction would contribute. But a Time by Subject interaction can be calculated *Within* each level of Program (i.e., $SS_{(T\times S)wP1}$, $SS_{(T\times S)wP2}$, $SS_{(T\times S)wP3}$), and aggregated to obtain $SS_{(T\times S)wP}$. $SS_{(T\times S)wP}$ would have df = (5-1)(3-1) + (5-1)(3-1) + (5-1)(3-1) = 24, and provides the error term for *time* and the *time* by *prog* interaction; that is, for effects that involve the Within-S *time* factor.

The data is entered into SPSS with a treatment code for the Between-S program factor, and three scores, one for each level of the Within-S time factor. An optional subject code is included below. The MANOVA and GLM commands combine Between-S (BY) and Within-S (/WSF) commands. EMMEANS is used to obtain means not provided by the default GLM but useful for analyzing numerators and interpreting results. A graph of the interaction is also requested (see Figure 9-1). The graph suggests that the results are consistent with expectation. Subjects in program 3 improved from time 1 to time 2 and maintained the improvement to time 3. The results for the other two groups vary and show no particular pattern across time. Averaged together they show a flat pattern, no change over Time. The predicted interaction is close to significant, p = .053.

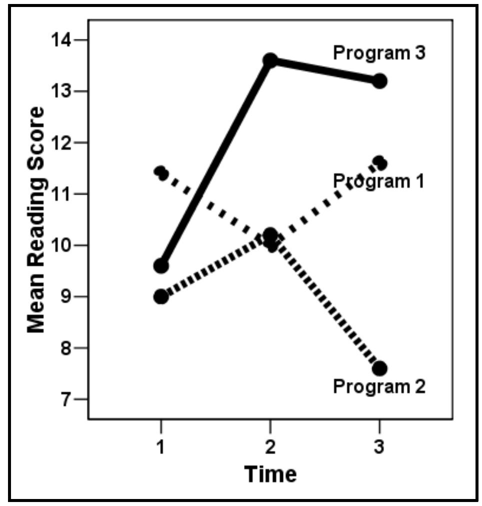


Figure 9-1. Graph of results

```
DATA LIST FREE / prog subj pre g1 g5.
BEGIN DATA
1 1 11  9  9     1 2 11 14 12     1 3  8  5 10     1 4 15 13 13     1 5 12  9 14
2 1 11 13  7     2 2 12 12  9     2 3  5  7  9     2 4  8  8  8     2 5  9 11  5
3 1 14 14 14     3 2  7 14 19     3 3 10 14  8     3 4  7 12 13     3 5 10 14 12
END DATA.

GLM pre g1 g5 BY prog /WSF = time(3) /PRINT = DESCR /PLOT = PROFILE(time*prog)
  /EMMEANS = TABLE(OVERALL)  /EMMEANS = TABLE(prog).
```

```
     prog  Mean    Std. Deviation N
 pre 1.00  11.4000 2.50998        5
     2.00  9.0000  2.73861        5
     3.00  9.6000  2.88097        5
     Total 10.0000 2.72554        15

 g1  1.00  10.0000 3.60555        5
     2.00  10.2000 2.58844        5
     3.00  13.6000 .89443         5
     Total 11.2667 2.96327        15
```

```
 g5  1.00  11.6000 2.07364        5
     2.00  7.6000  1.67332        5
     3.00  13.2000 3.96232        5
     Total 10.8000 3.52947        15

Tests of Within-Subjects Effects
 Source                      Type III SS     df     Mean Square F      Sig.
 time         Sphericity     12.311          2      6.156       1.106 .347
 time * prog  Sphericity     60.756          4      15.189      2.729 .053
 Error(time)  Sphericity     133.600         24     5.567               SS(T×S)wP

Tests of Between-Subjects Effects
 Source    Type III SS     df Mean Square F       Sig.
 Intercept 5141.356        1  5141.356    482.002 .000
 prog      78.978          2  39.489      3.702   .056
 Error     128.000         12 10.667                                    SSSwP

Estimated Marginal Means
1. Grand Mean
 Mean   Std.
        Error
 10.689 .487

2. prog
 prog Mean

 1.00 11.000
 2.00 8.933
 3.00 12.133
```

The plot and the ANOVA suggest that the reading program was successful. The *time* by *prog* interaction is marginally significant and could benefit from a more specific analysis. Also, the plot shows that the reading group performed better than the other groups in grades 1 and 5. Here is the equivalent MANOVA.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3).
```

```
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS             128.00      12     10.67
 prog                      78.98       2     39.49      3.70      .056

Tests involving 'TIME' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS             133.60      24      5.57
 TIME                      12.31       2      6.16      1.11      .347
 prog BY TIME              60.76       4     15.19      2.73      .053
```

Examining degrees of freedom helps to appreciate the nature of the error terms. Each level of *prog* has five subjects with $df = 5\text{-}1 = 4$ for SwP1, SwP2, and SwP3. Adding these together gives $df_{\text{Error}} = 3\times(5\text{-}1) = 12$ for SwP, as shown above for the Between-S error. Similarly, each level of *prog* has $df = (3\text{-}1)(5\text{-}1) = 8$ for (T×S)wP1, (T×S)wP2, and (T×S)wP3. Adding these gives $df_{\text{Error}} = 3\times(3\text{-}1)(5\text{-}1) = 24$ for (T×S)wP, as shown above for the Within-S error.

Conceptualizing the Error Terms

Although calculating SSs for the denominators manually would be a challenge, SPSS can produce quantities for single factor Within-S designs (i.e., $SS_{Subjects}$ and $SS_{Treatment×Subjects}$) at each level of *prog*, which can then be summed to obtain denominators for the mixed factorial. The ultimate values needed are SS_{SwP} = 128.0 (the denominator for *prog*) and $SS_{(T×S)wP}$ = 133.60 (the denominator for *time* and the *time* by *prog* interaction). Three single factor Within-S analyses, one for each level of *prog*, can be produced with the SPLIT FILE command, as below. Note that the *df*s for the denominators sum appropriately as do the SSs. If the *time* effect was partitioned for each single factor Within-S design, then the denominators would also be partitioned and they could be summed to obtain errors for a mixed factorial ANOVA in which *time* and *time* × *prog* effects were partitioned.

```
SPLIT FILE BY prog.
MANOVA pre g1 g5 /WSF = time(3).
prog:    1.00
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              66.67       4     16.67                       SSSwP1
 CONSTANT                1815.00       1   1815.00    108.90      .000
Tests involving 'TIME' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              27.73       8      3.47                       SS(T×S)wP1
 TIME                       7.60       2      3.80      1.10      .380

prog:    2.00
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              33.60       4      8.40                       SSSwP2
 CONSTANT                1197.07       1   1197.07    142.51      .000
Tests involving 'TIME' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              34.40       8      4.30                       SS(T×S)wP2
 TIME                      16.93       2      8.47      1.97      .202

prog:    3.00
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              27.73       4      6.93                       SSSwP3
 CONSTANT                2208.27       1   2208.27    318.50      .000
Tests involving 'TIME' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              71.47       8      8.93                       SS(T×S)wP3
 TIME                      48.53       2     24.27      2.72      .126
SPLIT FILE OFF.
```

$$SS_{SwP} = 128.0 = SS_{SwP1} + SS_{SwP2} + SS_{SwP3} = 66.67 + 33.60 + 27.73$$

$$df = 3×(5-1) = 12$$

$$SS_{(T×S)wP} = 133.60 = SS_{(T×S)wP1} + SS_{(T×S)wP2} + SS_{(T×S)wP3} = 27.73 + 34.40 + 71.47$$

$$df = 3×(3-1)(5-1) = 24$$

The nature of these error terms can also be shown with the data in Between-S format. Appendix 9-1 demonstrates with data in Between-S format that the error terms are indeed *subj WITHIN prog* and *time BY subj WITHIN prog*.

Planned Contrasts for Main Effects

The results can be analyzed further by planned contrasts for *time*, especially 1 versus 2&3, and *prog*, especially 1&2 versus 3, given the prediction that the reading readiness approach would be most effective. Numerators for the contrasts are calculated as usual. Cell means are used below to partition the main effects, rather than row or column means, and will be used later to generate contrasts to partition the interaction.

```
      T1P1  T1P2  T1P3  T2P1  T2P2  T2P3  T3P1  T3P2  T3P3
      11.4   9.0   9.6  10.0  10.2  13.6  11.6   7.6  13.2  L     SS = njL²/∑cj²
Time
 T1   −2    −2    −2     1     1     1     1     1     1     6.2  10.678 = 5×6.2²/18
 T2    0     0     0    −1    −1    −1     1     1     1    −1.4   1.633
                                                                  ∑ = SSTime
Program
 P1   −1    −1     2    −1    −1     2    −1    −1     2    13.0  46.944
 P2   −1     1     0    −1     1     0    −1     1     0    −6.2  32.033
                                                                  ∑ = SSProgram
```

The following MANOVA presents the results of these contrasts. The partitioning for the Within-S factor *time* is specified on the /WSD option, and the partitioning for the Between-S factor *prog* is specified on the /DESIGN option. The results include partitioned main effect numerators, as calculated above. Note that the denominator or error for the Between-S factor is not partitioned, whereas the error for the Within-S factor is. This parallels previous analyses for Within-S and Between-S factors. The primary interest in this study is the interaction, but focus first on the main effects shown in bold. The Program main effect, overall $p = .056$, did not benefit from the partitioning because $SS_{Program}$ for the numerator was divided close to equally between the two contrasts. The Time main effect, overall $p = .347$, did benefit because most of SS_{Time} loaded on the first contrast, df = 1, and the error was somewhat smaller, producing $p = .152$ for performance at time one before training compared to performance at times two and three after training.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3)
  /CONTR(time) = SPEC(1 1 1  -2  1 1   0 -1 1)
  /CONTR(prog) = SPEC(1 1 1  -1 -1 2  -1  1 0)
  /WSD time(1) time(2)  /DESIGN prog(1) prog(2).

Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL          128.00      12     10.67
 PROG(1)                   46.94       1     46.94      4.40      .058
 PROG(2)                   32.03       1     32.03      3.00      .109
...
```

```
Tests involving 'TIME(1)' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           54.80      12      4.57
 TIME(1)                   10.68       1     10.68      2.34      .152
 PROG(1) BY TIME(1)        38.27       1     38.27      8.38      .013
 PROG(2) BY TIME(1)          .42       1       .42       .09      .768
...
Tests involving 'TIME(2)' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           78.80      12      6.57
 TIME(2)                    1.63       1      1.63       .25      .627
 PROG(1) BY TIME(2)          .02       1       .02       .00      .961
 PROG(2) BY TIME(2)        22.05       1     22.05      3.36      .092
```

MANOVA also provides a detailed breakdown of main effects with the /PRINT = SIGNIFICANCE(...) option. Both SINGLEDF for Between-S factors and UNIVARIATE for Within-S factors are specified in the mixed factorial to partition both factors.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3)  /PRINT = SIGNIF(SINGLEDF UNIV)
  /CONTR(time) = SPECIAL(1 1 1  -2  1 1   0 -1 1)
  /CONTR(prog) = SPECIAL(1 1 1  -1 -1 2  -1  1 0).
```

```
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS             128.00      12     10.67
 prog                      78.98       2     39.49      3.70      .056
   1ST Parameter           46.94       1     46.94      4.40      .058
   2ND Parameter           32.03       1     32.03      3.00      .109
...
Tests involving 'TIME' Within-Subject Effect.
 EFFECT .. prog BY TIME
...
 EFFECT .. TIME
 Univariate F-tests with (1,12) D. F.
 Variable  Hypoth. SS   Error SS  Hypoth. MS  Error MS         F   Sig. of F
 T2          10.67778   54.80000    10.67778   4.56667  2.33820         .152
 T3           1.63333   78.80000     1.63333   6.56667   .24873         .627

 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS             133.60      24      5.57
 TIME                      12.31       2      6.16      1.11      .347
 prog BY TIME              60.76       4     15.19      2.73      .053
   1ST Parameter           38.29       2     19.14      3.44      .049
   2ND Parameter           22.47       2     11.23      2.02      .155
```

Special analyses for mixed designs can be more challenging with GLM than MANOVA. Below the MMATRIX and LMATRIX options are used to partition the main effects. The results agree with the earlier MANOVA results. Normalized coefficients can be used where necessary to produce correct SSs, as well as *F* and *p*, although the final statistics do not depend on the value of the coefficients.

```
*Partition main effect of Prog (Between-S factor).
GLM pre g1 g5 BY prog /WSF = time 3
  /LMATRIX prog -1 -1 2
  /LMATRIX prog -1  1 0.
...
```

```
Tests of Between-Subjects Effects
 Source    Type III SS     df Mean Square F       Sig.
 Intercept 5141.356        1  5141.356    482.002 .000
 prog      78.978          2  39.489      3.702   .056
 Error     128.000         12 10.667

Custom Hypothesis Tests #1
 L1       Contrast Estimate           7.506
          Std. Error                  3.578
          Sig.                        .058

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 46.944         1  46.944      4.401 .058
 Error    128.000        12 10.667

Custom Hypothesis Tests #2
 L1       Contrast Estimate           -3.580
          Std. Error                  2.066
          Sig.                        .109

 Source   Sum of Squares df Mean Square F     Sig.
 Contrast 32.033         1  32.033      3.003 .109
 Error    128.000        12 10.667
```

```
*Partition main effect of Time (Within-S factor) with normalized coefficients.
GLM pre g1 g5 BY prog /WSF = time(3)
  /MMATRIX pre -.8165 g1  .40825  g5 .40825;
           pre 0      g1 -.7071   g5 .7071.
```

```
...
Custom Hypothesis Tests
                                         T1              T2
 L1          Contrast Estimate           .844            -.330
             Std. Error                  .552            .662
             Sig.                        .152            .627
...
Univariate Test Results
 Source   Transformed     Sum of Squares df Mean Square F     Sig.
          Variable
 Contrast T1              10.678         1  10.678      2.338 .152
          T2              1.633          1  1.633       .249  .627

 Error    T1              54.800         12 4.567
          T2              78.798         12 6.567
```

Post Hoc Tests

Post hoc comparisons for mixed designs combine the Between-S and Within-S procedures. The POST HOC option can be used for the Between-S factor and EMMEANS for the Within-S factor. The Post Hoc menu in Figure 9-2 only lists the Between-S factor *prog*. The menu generated the following syntax, with default options removed.

None of the differences among the three *time* means are significant by the Bonferroni test, although time 1 versus time 3 would be by a liberal LSD test, $p = .127/3 = .042$.

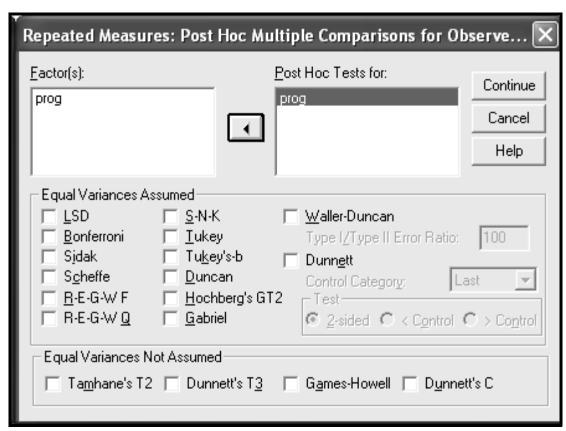


Figure 9-2. Post hoc tests for the Between-S factor

For the main effect of *prog*, groups 2 and 3 differ significantly by LSD (p = .020), SNK (different subsets), and Tukey (p = .049), but not by Bonferroni (p = .060 = 3 × .020).

```
GLM  pre g1 g5 BY prog  /WSFACTOR = time 3 Polynomial
  /POSTHOC = prog(LSD SNK TUKEY BONFERRONI)
  /EMMEANS = TABLES(time) COMPARE ADJ(BONF).
...
Estimated Marginal Means
Pairwise Comparisons
 (I)  (J)  Mean Difference Std.  Sig.(a)
 time time (I-J)           Error
 1    2    -1.267           .558   .127
      3    -.800           1.020  1.000
 2    3     .467            .936  1.000

Post Hoc Tests
prog
            (I)    (J)    Mean Difference Std.      Sig.
            prog   prog   (I-J)           Error
 LSD        1.0000 2.0000  2.066667       1.1925696 .109
                   3.0000 -1.133333       1.1925696 .361
            2.0000 3.0000 -3.200000(*)    1.1925696 .020

 Tukey HSD  1.0000 2.0000  2.066667       1.1925696 .233
                   3.0000 -1.133333       1.1925696 .620
            2.0000 3.0000 -3.200000(*)    1.1925696 .049

 Bonferroni 1.0000 2.0000  2.066667       1.1925696 .326
                   3.0000 -1.133333       1.1925696 1.000
            2.0000 3.0000 -3.200000       1.1925696 .060

Homogeneous Subsets
                 prog   N Subset
                          1         2
 Student-Newman- 2.0000 5  8.933333
 Keuls(a,b,c)    1.0000 5 11.000000 11.000000
                 3.0000 5           12.133333
                 Sig.     .109      .361

 Tukey           2.0000 5  8.933333
 HSD(a,b,c)      1.0000 5 11.000000 11.000000
                 3.0000 5           12.133333
                 Sig.     .233      .620
```

Partitioning the Interaction

Follow-up analyses for the interaction include partitioning the interaction and simple effects. To partition the interaction, pairs of coefficients for the main effect contrasts are multiplied to produce interaction contrasts. In the present study, the main effects of Time and Program both involve two contrasts. The following table reproduces the main effect contrasts from earlier and shows the products that partition the interaction.

	T1P1	T1P2	T1P3	T2P1	T2P2	T2P3	T3P1	T3P2	T3P3		
	11.4	9.0	9.6	10.0	10.2	13.6	11.6	7.6	13.2	L	$SS = n_j L^2/\sum c_j^2$
Time											
T1	−2	−2	−2	1	1	1	1	1	1		
T2	0	0	0	−1	−1	−1	1	1	1		
Program											
P1	−1	−1	2	−1	−1	2	−1	−1	2		
P2	−1	1	0	−1	1	0	−1	1	0		
T×P											
T1P1	2	2	−4	−1	−1	2	−1	−1	2	16.6	38.272
T2P1	0	0	0	1	1	−2	−1	−1	2	.2	.0167
T1P2	2	−2	0	−1	1	0	−1	1	0	1.0	.4167
T2P2	0	0	0	1	−1	0	−1	1	0	−4.2	22.05
											$\sum = SS_{T\times P}$

The following MANOVA is identical to one presented earlier and now shows results of contrasts for both interaction and main effects. As before, partitioning the Within-S factor *time* is specified on the /WSD option, and partitioning the Between-S factor *prog* is specified on the /DESIGN option. The results include partitioned interaction numerators, as calculated above. The error for the interaction is partitioned because it involves a Within-S factor, *time*.

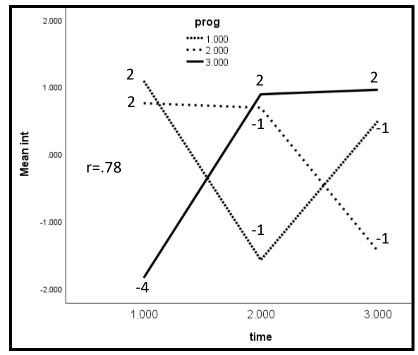


Figure 9-3. Interaction Deviations & Contrast Coefficients

Of primary interest in the present study is the first contrast for the interaction, calculated manually by multiplying the first contrast for *time* (Pre versus G1 & G5) by the first contrast for *prog* (P1 & P2 versus P3). Although the omnibus interaction effect was not quite significant, $p = .053$, this focussed contrast is significant, $p = .013$, because the contrast accounts for a substantial portion of $SS_{P\times T}$ (38.27 units of 60.76 units in total), involves a single *df*, and its error (4.57) is smaller than the error in the omnibus analysis (5.57). The contrast accounts for much of the variability because the contrast coefficients correlate well with the cell means, clearly when main effects are removed and groups 2 and 3 are combined. This relationship is shown in Figure 9-3. The interaction deviations are plotted and the contrast coefficients written by the corresponding cells, $r = .78 = \sqrt{(38.272/60.755)}$. The correspondence can be seen more clearly by averaging means for programs 1 and 2.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3)
  /CONTR(time) = SPEC(1 1 1  -2 1 1  0 -1 1)
  /CONTR(prog) = SPEC(1 1 1  -1 -1 2  -1 1 0)
  /WSD time(1) time(2)  /DESIGN prog(1) prog(2).
```

```
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL          128.00      12     10.67
 PROG(1)                   46.94       1     46.94      4.40      .058
 PROG(2)                   32.03       1     32.03      3.00      .109
...
```

```
Tests involving 'TIME(1)' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL            54.80     12      4.57
 TIME(1)                    10.68      1     10.68      2.34      .152
 PROG(1) BY TIME(1)         38.27      1     38.27      8.38      .013
 PROG(2) BY TIME(1)           .42      1       .42       .09      .768
...
Tests involving 'TIME(2)' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL            78.80     12      6.57
 TIME(2)                     1.63      1      1.63       .25      .627
 PROG(1) BY TIME(2)           .02      1       .02       .00      .961
 PROG(2) BY TIME(2)         22.05      1     22.05      3.36      .092
```

MANOVA also provides a detailed breakdown of the interaction with /PRINT = SIGNIFICANCE, as shown earlier.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3) /PRINT = SIGNIF(SINGLEDF UNIV)
  /CONTR(time) = SPECIAL(1 1 1  -2  1 1   0 -1 1)
  /CONTR(prog) = SPECIAL(1 1 1  -1 -1 2  -1  1 0).

Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              128.00     12     10.67
 prog                       78.98      2     39.49      3.70      .056
   1ST Parameter            46.94      1     46.94      4.40      .058
   2ND Parameter            32.03      1     32.03      3.00      .109
...
Tests involving 'TIME' Within-Subject Effect.
 EFFECT .. prog BY TIME
 EFFECT .. 1ST Parameter of prog BY TIME
 Univariate F-tests with (1,12) D. F.
 Variable   Hypoth. SS   Error SS    Hypoth. MS  Error MS          F  Sig. of F
 T2           38.27222   54.80000      38.27222   4.56667    8.38078       .013
 T3             .01667   78.80000        .01667   6.56667     .00254       .961

 EFFECT .. 2ND Parameter of prog BY TIME
 Univariate F-tests with (1,12) D. F.
 Variable  Hypoth. SS  Error SS  Hypoth. MS   Error MS         F   Sig. of F
 T2            .41667  54.80000      .41667    4.56667    .09124        .768
 T3          22.05000  78.80000    22.05000    6.56667   3.35787        .092

 EFFECT .. TIME
...
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              133.60     24      5.57
 TIME                       12.31      2      6.16      1.11      .347
 prog BY TIME               60.76      4     15.19      2.73      .053
   1ST Parameter            38.29      2     19.14      3.44      .049
   2ND Parameter            22.47      2     11.23      2.02      .155
```

When partitioning the interaction in GLM with MMATRIX and LMATRIX options, GLM automatically generates the interaction contrast(s). The first analysis below requests just the first contrast for the interaction and uses normalized coefficients to obtain the correct SS, which corresponds to our calculations and previous MANOVAs.

```
*Partition interaction in GLM, first contrast, normalized coefficients.

GLM pre g1 g5 BY prog /WSF = time(3)
  /MMATRIX pre  -.8165  g1 .40825  g5 .40825
  /LMATRIX prog -.40825   -.40825     .8165.
...
 L1       Contrast Estimate           2.767
          Std. Error                  .956
          Sig.                        .013

 Source   Sum of Squares df Mean Square F      Sig.
 Contrast 38.273         1  38.273      8.381 .013
 Error    54.800         12 4.567
```

Simple Effects of the Within-S Factor

The second follow-up test for interactions is a simple effects analysis. Numerator calculations for the mixed factorial are identical to other designs and shown below for the simple effect of the Within-S *time* factor within levels of the Between-S Program factor.

$\bar{y}_{pt}$		Time Pre	G1	G5	$\bar{y}_p$		
Prog	1	11.4	10.0	11.6	11.0000		
	$\bar{y}_{pt}-\bar{y}_{p1}$	.400	-1.000	.600		SS_{TwP1}=7.600	df = 2
	2	9.0	10.2	7.6	8.9333		
	$\bar{y}_{pt}-\bar{y}_{p2}$	.0667	1.2667	-1.3333		SS_{TwP2}=16.933	df = 2
	3	9.6	13.6	13.2	12.1333		
	$\bar{y}_{pt}-\bar{y}_{p3}$	-2.5333	1.4667	1.0667		SS_{TwP3}=48.533	df = 2

The simple effects analyses involve an alternative partitioning of SS_{Total}: $\sum SS_{TwP}$ = 7.600+16.933+48.533 = 73.066 = $SS_T + SS_{P\times T}$. Each of the three simple effects with *df* = 2 can be partitioned into *df* = 1 contrasts, as shown below for the reading preparation program. The contrast could also be done with all 9 cell means in a row and 0s for all cells except the three shown.

	Pre	G1	G5		
	9.6	13.6	13.2	L	SS
T1 vs 2&3	-2	1	1	7.6	48.133
T2 vs 3	0	-1	1	-.4	0.400
					$\sum$= 48.533 = SS_{TwP3}

$$SS_{L1wP3} = 5\text{x}7.6^2/6 = 48.133$$

Carrying out this simple effects analysis in MANOVA requires a new option because Within-S factors cannot occur on the /DESIGN option and Between-S factors cannot occur on the /WSD option. To overcome this limitation, MANOVA has a keyword MWITHIN that requests mixed within (i.e., simple) effects. MWITHIN can appear on the DESIGN or WSD options, depending on the desired simple effects analysis. Below, the Within-S factor *time* is stated on the /WSD option and then MWITHIN appears on /DESIGN before *prog(1)*, *prog(2)*, and *prog(3)*. MANOVA analyzes the simple effect of the Within-S factor at each level of the Between-S factor. The relevant sections of the output are printed in bold because

MANOVA's labelling is unclear. The relevant lines contain names for both a level of *time* and a level *prog*. SS_{Error} = 133.60 from the overall simple effect of the Within-S factor *time* is partitioned into separate error terms for follow-up analyses.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3)
  /CONTR(time) = SPEC(1 1 1  -2 1 1  0 -1 1)
  /CONTR(prog) = SPECIAL(1 1 1  -1 -1 2  -1 1 0)
  /WSD time             /DESIGN MWITHIN prog(1) MWITHIN prog(2) MWITHIN prog(3)
  /WSD time(1) time(2)  /DESIGN MWITHIN prog(1) MWITHIN prog(2) MWITHIN prog(3).

* * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   1 * * * * * * * *
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL          128.00      12     10.67
 MWITHIN PROG(1)         1815.00       1   1815.00    170.16      .000
 MWITHIN PROG(2)         1197.07       1   1197.07    112.23      .000
 MWITHIN PROG(3)         2208.27       1   2208.27    207.03      .000

Tests involving 'TIME' Within-Subject Effect.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL          133.60      24      5.57
 MWITHIN PROG(1) BY TIME    7.60       2      3.80       .68      .515
 MWITHIN PROG(2) BY TIME   16.93       2      8.47      1.52      .239
 MWITHIN PROG(3) BY TIME   48.53       2     24.27      4.36      .024  predicted

* * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   2 * * * * * * * *
Tests of Between-Subjects Effects.
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL          128.00      12     10.67
 MWITHIN PROG(1)         1815.00       1   1815.00    170.16      .000
 MWITHIN PROG(2)         1197.07       1   1197.07    112.23      .000
 MWITHIN PROG(3)         2208.27       1   2208.27    207.03      .000
...

Tests involving 'TIME(1)' Within-Subject Effect.
 Source of Variation                SS   DF        MS         F  Sig of F
 WITHIN+RESIDUAL                 54.80   12      4.57
 MWITHIN PROG(1) BY TIME(1)       1.20    1      1.20       .26      .618
 MWITHIN PROG(2) BY TIME(1)        .03    1       .03       .01      .933
 MWITHIN PROG(3) BY TIME(1)      48.13    1     48.13     10.54      .007 predicted
...
Tests involving 'TIME(2)' Within-Subject Effect.
 Source of Variation                SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL                 78.80      12      6.57
 MWITHIN PROG(1) BY TIME(2)       6.40       1      6.40       .97      .343
 MWITHIN PROG(2) BY TIME(2)      16.90       1     16.90      2.57      .135
 MWITHIN PROG(3) BY TIME(2)        .40       1       .40       .06      .809
```

The results are quite tidy. The overall effect of *time* is only significant for program 3, the reading-intensive program, and does not approach significance for the other two groups. Also, the simple effect of *time* for program 3 is entirely accounted for by the difference between pre-test reading scores and scores in Grades 1 and 5, as predicted if the program effect does not fade. No other effects approach significance.

Chapters 7 and 8 showed that Within-S contrasts can be calculated by creating individual contrast scores for each subject and then using the mean and standard deviation of the contrast scores to test H_0:

$\mu_{Contrast} = 0$. The complication for the mixed design is that the Within-S factor is nested within a Between-S factor. The appropriate analysis in terms of contrast scores requires aggregating over the levels of the Between-S factor, specifically the error terms. The following COMPUTE calculates normalized contrast scores and the SPLIT file produces three MANOVAs, one for each level of *prog*.

```
COMPUTE t1 = (-2*pre+1*g1+1*g5)/SQRT(6).
SPLIT FILE BY prog.
MANOVA t1.

prog:      1.00
 Tests of Significance for t1 using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS               7.13       4      1.78                       (T×S)wP1
 CONSTANT                   1.20       1      1.20       .67      .458

prog:      2.00
 Tests of Significance for t1 using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS               8.80       4      2.20                       (T×S)wP2
 CONSTANT                    .03       1       .03       .02      .908

prog:      3.00
 Tests of Significance for t1 using UNIQUE sums of squares
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN CELLS              38.87       4      9.72                       (TxS)wP3
 CONSTANT                  48.13       1     48.13      4.95      .090

SPLIT FILE OFF.
```

Note that the SSs in bold agree with the numerators in earlier simple effects analyses, but the denominators do not. The three denominators must be summed to obtain the appropriate error term. One way to do that is to request the simple effects of t1 within the levels of Program. Now the results agree with the earlier analyses. $SS_{Error} = 54.80 = 7.13+8.80+38.87$, the separate errors from the previous analysis.

```
MANOVA t1 BY prog (1 3)
   /DESIGN MWITHIN prog(1) MWITHIN prog(2) MWITHIN prog(3).

 Tests of Significance for t1 using UNIQUE sums of squares
 Source of Variation      SS  DF     MS         F  Sig of F
 WITHIN+RESIDUAL       54.80  12   4.57                        (T×S)wP =7.13+8.80+38.87
 MWITHIN PROG(1)        1.20   1   1.20       .26      .618
 MWITHIN PROG(2)         .03   1    .03       .01      .933
 MWITHIN PROG(3)       48.13   1  48.13     10.54      .007
```

Using EMMEANS to request the simple effects of the Within-S factor Time in GLM performs pairwise comparisons, but does not produce the overall simple effects tests; multivariate statistics are reported for overall simple effects, as shown earlier for the Within-S factorial. By these tests, the simple effect of *time* is significant for program 3 but not for programs 1 and 2, agreeing with the MANOVA results. The pairwise comparisons do lead to a meaningful pattern: none of the differences are significant for programs 1 and 2, whereas time 1 differs from each of times 2 and 3 (marginally by non-directional tests) for program 3. This pattern makes sense theoretically.

```
GLM pre g1 g5 BY prog /WSF = time 3 SPECIAL(1 1 1  -2 1 1  0 -1 1)
  /CONTRAST(prog) = SPEC(-1 -1 2  -1 1 0)
  /EMMEANS = TABLE(prog BY time) COMPARE (time).
...
Estimated Marginal Means
Pairwise Comparisons
 prog    (I)  (J)  Mean Difference Std.  Sig.(a)
        time time (I-J)           Error

 1.0000 1    2    1.400           .966  .173
             3    -.200           1.766 .912
        2    3    -1.600          1.621 .343

 2.0000 1    2    -1.200          .966  .238
             3    1.400           1.766 .443
        2    3    2.600           1.621 .135

 3.0000 1    2    -4.000(*)       .966  .001
             3    -3.600          1.766 .064
        2    3    .400            1.621 .809
Multivariate Tests
 prog                   Value F        Hypothesis df Error df Sig.
 1.0000 Pillai's trace  .185  1.250(a) 2.000         11.000   .324
...
 2.0000 Pillai's trace  .232  1.663(a) 2.000         11.000   .234
...
 3.0000 Pillai's trace  .590  7.910(a) 2.000         11.000   .007
...
```

Simple Effects of the Between-S Factor

The previous analyses tested the simple effects of the Within-S factor *time* within levels of the Between-S factor *prog*. The following analyses test the simple effects of the Between-S factor *prog* within levels of the Within-S factor *time*. Numerators for overall and partitioned simple effects are calculated as usual.

	Program 1	2	3	$\bar{Y}_{t2}$		
$\bar{Y}_{pt2}$	10.00	10.20	13.60	11.267		
$\bar{Y}_{pt2}-\bar{Y}_{t2}$	-1.267	-1.067	2.33			
	$SS_{PwT2} = 5\times(-1.267^2 + -1.067^2 + 2.33^2) = 40.926$					
				L	SS	
P1 = P12v3	-1	-1	2	7.0	40.833	$= SS_{P1wT2}$
P2 = P1v2	-1	1	0	0.2	0.10	$= SS_{P2wT2}$
				Sum =	40.933	$= SS_{PwT2}$

Corresponding SPSS analyses follow. Note the differences between the analysis here and the earlier simple effects with respect to error terms. Specifically, error terms from the overall simple effects are not partitioned for follow-up analyses when the Between-S factor *prog* is partitioned.

```
MANOVA pre g1 g5 BY prog(1 3) /WSF = time(3)
  /CONTR(time) = SPECIAL(1 1 1  -2  1 1   0 -1 1)
  /CONTR(prog) = SPECIAL(1 1 1  -1 -1 2  -1  1 0)
  /WSD MWITHIN time(1) MWITHIN time(2) MWITHIN time(3) /DESIGN prog
  /WSD MWITHIN time(1) MWITHIN time(2) MWITHIN time(3) /DESIGN prog(1) prog(2).

* * * * * * A n a l y s i s   o f   V a r i a n c e -- design   1 * * * * * *
Tests involving 'MWITHIN TIME(1)' Within-Subject Effect.
 Source of Variation              SS    DF        MS          F  Sig of F
 WITHIN+RESIDUAL               88.40    12      7.37
 MWITHIN TIME(1)             1500.00     1   1500.00     203.62      .000
 PROG BY MWITHIN TIME (1)      15.60     2      7.80       1.06      .377   no effect
...
Tests involving 'MWITHIN TIME(2)' Within-Subject Effect.
 Source of Variation              SS    DF        MS          F  Sig of F
 WITHIN+RESIDUAL               82.00    12      6.83
 MWITHIN TIME(2)             1904.07     1   1904.07     278.64      .000
 PROG BY MWITHIN TIME (2)      40.93     2     20.47       3.00      .088   predicted
...
Tests involving 'MWITHIN TIME(3)' Within-Subject Effect.
 Source of Variation              SS    DF        MS          F  Sig of F
 WITHIN+RESIDUAL               91.20    12      7.60
 MWITHIN TIME(3)             1749.60     1   1749.60     230.21      .000
 PROG BY MWITHIN TIME (3)      83.20     2     41.60       5.47      .020   predicted

* * * * * * A n a l y s i s   o f   V a r i a n c e -- design   2 * * * * * *
Tests involving 'MWITHIN TIME(1)' Within-Subject Effect.
 Source of Variation              SS    DF        MS          F  Sig of F
 WITHIN+RESIDUAL               88.40    12      7.37                    same error
 MWITHIN TIME(1)             1500.00     1   1500.00     203.62      .000
 PROG(1) BY MWITHIN TIME (1)    1.20     1      1.20        .16      .694 no effect
 PROG(2) BY MWITHIN TIME (2)   14.40     1     14.40       1.95      .187

Tests involving 'MWITHIN TIME(2)' Within-Subject Effect.
 Source of Variation              SS    DF        MS          F  Sig of F
 WITHIN+RESIDUAL               82.00    12      6.83                    same error
 MWITHIN TIME(2)             1904.07     1   1904.07     278.64      .000
 PROG(1) BY MWITHIN TIME (2)   40.83     1     40.83       5.98      .031 predicted
 PROG(2) BY MWITHIN TIME (2)     .10     1       .10        .01      .906

Tests involving 'MWITHIN TIME(3)' Within-Subject Effect.
 Source of Variation              SS    DF        MS          F  Sig of F
 WITHIN+RESIDUAL               91.20    12      7.60                    same error
 MWITHIN TIME(3)             1749.60     1   1749.60     230.21      .000
 PROG(1) BY MWITHIN TIME (3)   43.20     1     43.20       5.68      .035 predicted
 PROG(2) BY MWITHIN TIME (3)   40.00     1     40.00       5.26      .041
```

These analyses are actually equivalent to separate single factor Between-S analyses of variance, one for each level of the *time* factor, as shown below for the time 3 measure *g5*.

```
*Prog within Time(3).
MANOVA g5 BY prog(1 3)
  /CONTR(prog) = SPECIAL(1 1 1  -1 -1 2  -1 1 0)  /DESIGN  /DESIGN prog(1) prog(2).
* * * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   1 * * * * * * * *
 Source of Variation           SS       DF        MS          F  Sig of F
 WITHIN CELLS               91.20       12      7.60
 prog                       83.20        2     41.60       5.47      .020
```

```
* * * * * * * * A n a l y s i s   o f   V a r i a n c e -- Design   2 * * * * * * *
 Source of Variation          SS      DF        MS         F  Sig of F
 WITHIN+RESIDUAL           91.20      12      7.60
 PROG(1)                   43.20       1     43.20      5.68      .035
 PROG(2)                   40.00       1     40.00      5.26      .041
```

Overall, this simple effects analysis is readily interpreted in terms of the hypotheses. The program effect is marginally significant at time 2, but the difference between program 3 and programs 1 and 2 is significant, as predicted. At time 3, the overall effect of *prog* is significant, as is the difference between program 3 and programs 1 and 2. A small complication at time 3 is that the difference between programs 1 and 2 is significant; the social control group performs more poorly than the no-treatment control group.

Since the GLM /EMMEANS option allows for both Within-S and Between-S factors, it appears straightforward to obtain the overall simple effects, and indeed this is true for the simple effect of the Between-S factor *prog* as shown below. This analysis could also be done using separate analyses for each level of the Within-S factor *time*, as just illustrated for MANOVA. Rather than partitioning the simple effect as predicted, however, EMMEANS produces pairwise comparisons, which can lead to awkward results. In the first analysis, for example, program 3 differs from programs 1 and 2 at time 2, a meaningful pattern, but programs 1 and 2 differ significantly at time 3, as do programs 2 and 3, but programs 1 and 3 do not. This pattern is not easy to explain theoretically.

```
GLM pre g1 g5 BY prog /WSF = time 3 SPECIAL(1 1 1  -2 1 1  0 -1 1)
  /CONTRAST(prog) = SPEC(-1 -1 2  -1 1 0)
  /EMMEANS = TABLE (prog BY time) COMPARE (prog).
...
Estimated Marginal Means
Pairwise Comparisons
 time (I)    (J)    Mean Difference Std.  Sig.(a)
      prog   prog   (I-J)           Error
 1    1.0000 2.0000 2.400           1.717 .187
             3.0000 1.800           1.717 .315
      2.0000 3.0000 -.600           1.717 .733

 2    1.0000 2.0000 -.200           1.653 .906
             3.0000 -3.600          1.653 .050
      2.0000 3.0000 -3.400          1.653 .062

 3    1.0000 2.0000 4.000(*)        1.744 .041
             3.0000 -1.600          1.744 .377
      2.0000 3.0000 -5.600(*)       1.744 .007

Univariate Tests
 time          Sum of Squares df Mean Square F     Sig.
 1    Contrast 15.600         2  7.800       1.059 .377
      Error    88.400         12 7.367

 2    Contrast 40.933         2  20.467      2.995 .088
      Error    82.000         12 6.833

 3    Contrast 83.200         2  41.600      5.474 .020
      Error    91.200         12 7.600
```

Given the limitations of GLM, MANOVA is often preferred to carry out simple effects analyses for mixed factorials.

Conclusions

We have now largely completed work on analysis of variance, having covered single factor and factorial designs involving Within-S, Between-S, or a mix of Between-S and Within-S factors. One of the admirable features of sophisticated programs like MANOVA and GLM is that they automatically determine the appropriate error terms for different designs. As a result, the principles behind error terms for Between-S and Within-S designs, generalize to studies involving multiple factors with various combinations of Within-S and Between-S factors. Appendix 9-2 illustrates with a three factor ANOVA involving one Between-S factor and two Within-S factors.

Appendix 9-1: Mixed Analysis in Between-S Format

Entering the data in Between-S format (i.e., one observation per row with a Subject factor added) demonstrates the various error terms used in preceding analyses and allows a comparison between the Within-S and Between-S analyses, specifically the possible benefits of a smaller error term with the latter.

```
*Between-S format.
DATA LIST FREE / prog subj time score.
BEGIN DATA
1 1 1 11  1 1 2  9  1 1 3  9        1 2 1 11  1 2 2 14  1 2 3 12
1 3 1  8  1 3 2  5  1 3 3 10        1 4 1 15  1 4 2 13  1 4 3 13        1 5 1 12  1 5 2  9  1 5 3 14
2 1 1 11  2 1 2 13  2 1 3  7        2 2 1 12  2 2 2 12  2 2 3  9
2 3 1  5  2 3 2  7  2 3 3  9        2 4 1  8  2 4 2  8  2 4 3  8        2 5 1  9  2 5 2 11  2 5 3  5
3 1 1 14  3 1 2 14  3 1 3 14        3 2 1  7  3 2 2 14  3 2 3 19
3 3 1 10  3 3 2 14  3 3 3  8        3 4 1  7  3 4 2 12  3 4 3 13        3 5 1 10  3 5 2 14  3 5 3 12
END DATA.
```

The following analysis shows the sources of variability for the mixed design: the main effect of *prog*, the main effect of *time*, the *prog by time* interaction, the *subj within prog* error term for the Between-S effect of prog, and the *time by subj within prog* error term for the main and interaction effects involving the Within-S effect of time. The values here correspond to those in the initial default ANOVA.

```
MANOVA score BY prog(1 3) time(1 3) subj(1 5)
  /DESIGN prog   time   prog BY time   subj WITHIN prog   time BY subj WITHIN prog.
```

Source of Variation	SS	DF	MS	F	Sig of F
RESIDUAL	.00	0	.		
PROG	78.98	2	39.49	.	.
TIME	12.31	2	6.16	.	.
PROG BY TIME	60.76	4	15.19	.	.
SUBJ WITHIN PROG	128.00	12	10.67	.	.
TIME BY SUBJ WITHIN PROG	133.60	24	5.57	.	.

Given the Between-S format, we can also determine the results for a Between-S factorial ignoring the fact that *time* is a Within-S factor. The analysis below treats both *time* and *prog* as Between-S factors. The *prog* by *time* interaction is more significant in the Within-S design (p = .053) despite numerators being the same because MS_{Error} for the Within-S design is smaller (5.567) than for the Between-S design (7.27). An advantage of Within-S designs is smaller error terms and increased sensitivity to effects.

```
*Default Between-S factorial.
MANOVA score BY prog(1 3) time(1 3).
```

Source of Variation	SS	DF	MS	F	Sig of F	
WITHIN CELLS	261.60	36	7.27			
prog	78.98	2	39.49	5.43	.009	
time	12.31	2	6.16	.85	.437	
prog BY time	**60.76**	**4**	**15.19**	**2.09**	**.102**	***Not significant***
(Total)	413.64	44	9.40			

Given data in Between-S format, MANOVA allows users to identify denominators and associate them with the appropriate numerators, as shown below. The numbers 1 and 2 identify the denominators. SwP is the appropriate error for the Between-S effect of program, and T×SwP is the appropriate error term for factors that include the Within-S time factor, namely the main effect of Time and the Time by Program interaction.

```
MANOVA score BY prog(1 3) time(1 3) subj (1 5)
  /DESIGN prog vs 1 time vs 2 prog BY time vs 2
  subj w prog  = 1  time BY subj W prog = 2.
```

Source of Variation	SS	DF	MS	F	Sig of F
Error 1	128.00	12	10.67		
PROG	78.98	2	39.49	3.70	.056
Error 2	133.60	24	5.57		
TIME	12.31	2	6.16	1.11	.347
PROG BY TIME	60.76	4	15.19	2.73	.053

Appendix 9-2: Generalization to Higher Order Designs

Researchers often use ANOVA to examine the influence of three or more factors. The SPSS commands, output, and interpretation are generalizations of the elements discussed here. To illustrate, consider a three-factor design with 2 levels of A, 3 levels of B, 4 levels of C, and 5 levels of S (subjects). Factor A is Between-S and factors B and C are Within-S factors. Here is a hypothetical dataset. Each of 5 subjects in the two levels of A has 3×4 = 12 scores.

a	s	b1c1	b1c2	b1c3	b1c4	b2c1	b2c2	b2c3	b2c4	b3c1	b3c2	b3c3	b3c4
1	1	95	102	87	95	104	98	103	105	74	96	102	101
1	2	109	84	98	103	104	114	88	106	102	100	115	99
1	3	95	85	103	69	104	105	106	92	98	88	87	82
1	4	92	95	101	107	110	95	88	108	110	85	107	99
1	5	97	103	114	86	114	112	120	106	95	110	124	81
2	1	106	88	92	100	101	97	87	94	112	101	122	96
2	2	89	92	110	115	101	93	107	96	87	104	90	106
2	3	102	112	99	110	105	93	107	108	100	100	97	114
2	4	96	105	95	97	113	97	116	100	94	99	121	105
2	5	109	112	109	105	102	111	95	101	106	118	110	104

The SPSS command to analyze the data uses BY for the Between-S factor A and /WSF for the Within-S factors B and C. Note that b(3) occurs first and c(4) second because of the order that the 12 variables are listed on GLM. Given variables were listed on GLM in the order: b1c1 b2c1 b3c1 ...b1c4 b2c4 b3c4, /WSF = c(4) b(3) would be correct.

```
GLM b1c1 b1c2 b1c3 b1c4 b2c1 b2c2 b2c3 b2c4 b3c1 b3c2 b3c3 b3c4 BY a /WSF = b(3) c(4).
```

Tests of Within-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
b	256.317	2	128.158	1.961	.173
b * a	571.217	2	285.608	4.371	.031
Error(b)	1045.467	16	65.342		
c	259.567	3	86.522	.666	.581
c * a	201.200	3	67.067	.516	.675
Error(c)	3118.067	24	129.919		
b * c	506.283	6	84.381	1.175	.335
b * c * a	326.450	6	54.408	.758	.607
Error(b*c)	3446.933	48	71.811		

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1222100.833	1	1222100.833	6301.384	.000
a	320.133	1	320.133	1.651	.235
Error	1551.533	8	193.942		

The partitioning of SS_{Total} into 11 components is illustrated by the solid lines in the following tree diagram. To summarize,

$$SS_{Total} = SS_A + SS_B + SS_C + SS_{AB} + SS_{AC} + SS_{BC} + SS_{ABC} + SS_{SwA} + SS_{B\times SwA} + SS_{C\times SwA} + SS_{B\times C\times SwA}$$

There are seven effects of interest, each giving a unique $SS_{Numerator}$: A, B, C, AB, AC, BC, and ABC.

In a completely Between-S design there would be a single error term (denominator) based on the variability among scores in each of the eight conditions (cells): SS_{SwABC}. In a completely Within-S design, there would be variability due to subjects (SS_S) and seven unique error terms, one for each effect defined in terms of interactions with subjects (e.g., $SS_{A\times S}$, $SS_{A\times B\times S}$). In the mixed design here with one Between-S factor and two Within-S factors, there are four error terms. The error term for the Between-S factor is SS_{SwA}, the variability in subjects within levels of A. The error term for Within-S factors and interactions involve interactions with subjects, but now nested within levels of A (e.g., $SS_{B\times SwA}$). Dashed lines show which denominator goes with which of the main effects, two-way interactions, and three-way interaction. Given the tree and the output above, calculate df for each of the components.