

Analysis of Designs with Detached Control Groups
Robert M. Hamer, Virginia Commonwealth University

Introduction

This paper assumes that the reader understands (1) the general linear model, and its parameters as defined in the usual fixed effects ANOVA models, (2) hypothesis testing in such models using linear combinations of the parameters, and (3) cell means parameterizations of such models. It further assumes that the reader is familiar with the use of the CONTRAST statement in PROC GLM to test linear combinations of cell means.

Sometimes it makes sense to use a factorial design but add one or more additional control groups to the design. You may have several factors, each with several levels, but there is no level on a factor corresponding to the absence of treatment on that factor, or perhaps there no cell corresponding to the absence of treatment on all factors. In such situations, we might wish to add one or more detached control groups. A detached control group is an additional cell, not part of a factorial design. You are interested in comparing this group or groups to the groups in the factorial design.

For example, suppose that an archaeologist is interested in the effect of tempering material on pottery. Tempering material is a substance that is added to the clay from which pottery is made, usually for religious reasons. An archaeologist may wish to assess the changes in physical characteristics of the pottery that the tempering material produces. In the Southeastern United States, three types of tempering materials were often added to clay when making pottery from it: sand, crushed shells, and crushed shells that had been burned. Similarly, in the Southeastern United States, tempering materials were usually added to the clay mixture in two amounts: corresponding to about twenty and forty percent.

Experimental Design

The archaeologist was interested in demonstrating that although the type and amount of tempering material was usually added for religious reasons, the use of such tempering material generally had beneficial effects on the pottery mixture. To test this hypothesis he designed a 3 x 2 completely randomized factorial (CRD) (see Figure 1) design with two factors: tempering material (sand, unburned shell, and burned shell) and two amounts (twenty and forty percent). Thus, each cell represents one of three materials added in one of two amounts. Under each condition he made four briquettes with a combination of a particular tempering material at a particular amount and fired them at a 1000 degrees centigrade, which was the approximate

temperature at which such pottery was usually fired. Note that none of the six resulting cells included briquettes made with no tempering material, that is, pure clay.

The experimenter wished further to demonstrate that adding tempering material to clay has a nonnull effect. Thus, an additional four briquettes containing no tempering material whatsoever were made. These four briquettes represent a control group for the 2 x 3 design, but they are not part of the design. This group of experimental units is called a detached control group. Additionally, in an attempt to discover the effect of temperature on clay, the archaeologist made two more sets of four briquettes of pure clay: one which was fired at 800 degrees and one which was fired at 1200 degrees. Thus, if we were to consider the three groups of clay without any tempering material, they represent a three-group one-way CRD in which the single factor (temperature) has three levels.

Figure 1. Schematic diagram of 3 x 2 factorial design with 3 detached control groups, and a one-way design in the control groups.

Factorial Portion

Material	Amount			
	20 Percent		40 Percent	
	CRACK	SHATTER	CRACK	SHATTER
Sand	36.00	16.42	...	
	.	.		
	.	.		
UBS				
BS				

Control Portion

CRACK	SHATTER	Temperature		
		800	1000	1200
		CRACK	SHATTER	CRACK
26.80	12.82
.	.			
.	.			
.	.			

Two dependent variables were measured: the amount of force required to put a visible crack in a briquette (CRACK), and the amount of additional force required to break the briquette into two or more separate pieces (SHATTER). (This design was adapted from a study done by Bronitsky and Hamer, *American Antiquity*, 1986, 89-101.)

Analysis Strategies

There are several ways in which we might propose to analyze these data, each with advantages and disadvantages:

1. Do a 3 x 2 factorial MANOVA and a separate three group one-way MANOVA.
2. Subtract the mean of the control groups from all the data in the factorial part of the design and analyze that part alone.
3. Fit a one-way nine group CRD cell means model, and test hypotheses corresponding to the main effects and interactions from a factorial analysis, and corresponding to the "treatment" effect in those cells which vary on temperature using contrast statements.

We will discuss the advantages of each of these in turn.

1. Do a 3 x 2 factorial MANOVA and a separate three group one-way MANOVA. This proposed analysis simply separates the data into two datasets: one from the 3 x 2 factorial MANOVA part, and one from the three group one-way MANOVA part. Observations in each part are treated separately, and analyzed separately. The advantage to this is that it is simple, and not obviously incorrect, but there are sufficient disadvantages so that it should not be done.

First, by separating the data into two sets of observations, and analyzing them as two separate experiments, we lose the ability to compare the groups that got tempering material with the groups that did not. In fact, we lose the ability to make any comparisons between any of the groups (or linear combination of them) that got tempering material, and any of the groups (or linear combination of them) that did not get tempering material. Presumably, at least part of the reason that the study was designed with both experimental and control groups was to be able to compare them.

Second, by separating the data into two sets of observations, and analyzing them as two separate experiments, we lose degrees of freedom (df), in that the error df are based upon the number of observations in the analysis, minus the model df. By using only some of the observations in an analysis, we decrease the error

df. Since the mean square error (MSE) consists of the error sums of square divided by the error df, a smaller error df in the denominator means a larger error MSE. Since the MSE is the denominator in all the F-tests in the analysis, a larger MSE makes it more difficult to reject the null hypothesis for a given mean square, model (MSM). Thus, using only part of the observations in an analysis is less powerful, and less efficient, than using all of the observations, if it is possible to use all of the observations.

Third, one assumption of ANOVA is that the population MSE is the same, whether we are talking about the portion of the data that come from the factorial part or the one-way part of the study. If so, then an estimate of the MSE that did not involve all the observations would be a less precise estimate than one that did, if it is possible to estimate the MSE using all the observations.

2. Subtract the mean of the control groups from all the data in the factorial part of the design and analyze that part alone. This addresses the first objection above (the inability to do a comparison between observations that received tempering material and those that did not), but does so in both an inefficient and wrong manner. Since the hypothesis tests in a factorial ANOVA are invariant with respect to the addition of a constant, this would not change any of the hypothesis tests of the main effects or interactions. It would change the test on the constant, or intercept, making that test have meaning. The meaning would be that the constant would then be an estimate of the difference between the mean of the control groups and the mean of the experimental groups, and the test of whether the constant was zero would be the test of the difference between the experimental and control groups.

However, that approach is inefficient because the observations used in the control groups would not contribute to the error df, and hence any hypothesis tests are less efficient than would be hypothesis tests that did use those df. It would also be wrong, because we would be treating the mean of the control group or groups as a constant with no variance, rather than as a sample statistic. For the same reason, we would consider it improper to test the difference between the means of two groups by subtracting the mean of one group from all the observations in the other group, and then doing a one-group t- test.

3. Fit a one-way nine-group CRD and test hypotheses corresponding to the ones we could test in the analysis of the factorial design and in the analysis of the one-way design using CON-

TRAST statements in GLM. This analysis uses all the observations in the estimation of its MSE and hence is more efficient and precise than the analyses proposed above.

In the remainder of this paper, I will (1) create a SAS dataset, (2) present a preliminary analysis, and (3) two of the three proposed analyses using a cell means model, with contrasts between the parameters of the model, (4) discuss the contrasts in detail, and (5) present an optimal analysis. I will not present the second proposed analysis (subtraction of the mean of the control groups from the factorial-group data and analysis of those data) as I think that analysis is silly.

Creation of SAS Dataset

Figure 2. Creation of SAS dataset.

```
data full;
  input material $ 1-5 amount 7-8 temp 10-14
        (crack shatter) (016 6.2 6.2) group 30;
  cards;
[data lines here]
;
```

Figure 2 contains a data step used to create the SAS data set which we will use in this analysis. You will notice that we have a variable called MATERIAL that can have values of SAND, UBS and BS, a variable called AMOUNT which can contain values of 20 or 40, a variable called TEMP (for temperature) which is missing for the cells of the design corresponding to the factorial design and has values of either 800, 1000, or 1200 for the portion corresponding to the one-way design. The two dependent variables are CRACK and SHATTER, and finally there is a variable called GROUP with values from one to nine. Thus, the variable called GROUP simply defines the nine cells in this design without regard to their structure.

Preliminary Analysis

In the preliminary analysis in which we simply check for the equality of the group mean vectors among the nine groups (Figure 3). However, even if we find a difference, that does not tell us which groups differ from which others; it does not tell us on which variables these differences occur, and it tells us nothing about the effects of tempering material, amount, firing temperature, or any interactions. (Although I have not included any computer output from this program, there was indeed a significant GROUP effect telling us that there were differences among the groups.)

Figure 3. Preliminary analysis of differences among 9 means.

```
proc glm data=full;
  class group;
  model crack shatter = group;
  means group;
  manova h=group;
  title 'Preliminary simple MANOVA';
run;
```

Proposed Separate Datasets Analysis 1

The first plausible analysis was to create two separate SAS data sets, one containing the factorial part of the design (the cells which got tempering material of some sort and in some amount) and one containing the control part of the design (the cells which got no tempering material and thus contained pure clay but were fired at three different temperatures). Figure 4 contains a small SAS program to form the two datasets. With the twenty-four observations from the factorial portion of the design in a separate SAS data set, we may analyze them as a standard 3 x 2 factorial.

Figure 4. SAS Program to split 36 observation SAS dataset into a 24 observation SAS dataset and a 12 observation SAS dataset.

```
data factor control;
  set full
  if temp=. then output factor;
  else output control;
```

Figure 5 contains the SAS program to perform the factorial MANOVA, and Figure 6 contains the univariate output for the dependent variable CRACK. The R-square was about .53 (not shown).

Figure 5. SAS program to analyse a 3 x 2 factorial MANOVA.

```
proc glm data=factor;
  class material amount;
  model crack shatter
    = material|amount;
  manova h=material amount
    material*amount;
  title 'Two-Factor MANOVA';
run;
```

Examination of the F statistics for the materials effect, the amount effect, and the material by amount interaction reveal what appears to be a statistically significant difference between the two amounts: twenty percent verses forty percent.

Figure 6. Univariate output for CRACK.

Source	DF	Type III SS	F Value	Pr > F
MATERIAL	2	128.52503333	1.17	0.3331
AMOUNT	1	922.68400417	16.79	0.0007
MATERIAL*AMOUNT	2	26.42773333	0.24	0.7888

We can perform exactly the same analysis by using a cell means model (that is fitting a design in which the parameters of the design are the means of the six cells) and then using contrasts to construct our effects as linear combinations of these cell means. Figure 7 contains the program to perform this analysis.

Figure 7. SAS program to fit a cell means model to the 3 x 2 factorial data, and test main effects and interactions using contrasts.

```
proc glm data=factor;
  class group;
  model crack shatter = group/solution point;
  means group;
  contrast 'Material'
    group .5 .5 -.5 -.5 0 0,
    group 0 0 .5 .5 -.5 -.5;
  contrast 'Amount'
    group .33333 -.33333 .33333 -.33333
    .33333 -.33333;
  contrast 'Material*Amount'
    group .5 -.5 -.5 .5 0 0,
    group 0 0 .5 -.5 -.5 .5;
  manova h=group;
  title 'Cell means factorial';
run;
```

If you examine the contrasts carefully, you will notice that the contrast labelled MATERIAL is a 2 degree-of-freedom contrast in which the first one degree-of-freedom row represents the difference between the first level of MATERIAL and the second level and the second one degree-of-freedom row represents the difference between the second and the third levels. The first row of this contrast (.5 .5 -.5 -.5 0 0) forms a linear combination which is the mean of the first and second cells (those that used SAND) minus the mean of the third and fourth cells (those that used UBS). The second row of this contrast (0 0 .5 .5 -.5 -.5) forms a linear combination which is the mean of the third and fourth cells (those that used UBS) minus the mean of the fifth and sixth cells (those that received BS). If both differences are simultaneously zero, then the three marginal means are equal, and there is no difference among the three materials. The two rows of this contrast matrix are linearly independent, thus fulfilling the requirements for a contrast to test this three level effect.

Examination of the second contrast, labelled AMOUNT, shows that it consists of the average of the three cells that got forty percent tempering material vs. the three cells that got twenty percent tempering material. This contrast (.33333 -.33333 .33333 -.33333 .33333 -.33333) forms a linear combination consisting of the mean of cells 1, 3, and 5 (those that received 20 percent tempering material) minus the mean of cells 2, 4, and 6 (those that received 40 percent tempering

material). If this contrast is not significantly different from zero, it means that the cells receiving 20 percent tempering material were not significantly different from those receiving 40 percent tempering material. The contrast labelled the MATERIAL by AMOUNT interaction was obtained as the horizontal direct product of the two main effect contrasts involved.

Figure 8. Univariate output for CRACK

Contrast	DF	Contrast SS	F Value	Pr > F
Material	2	128.52503333	1.17	0.3331
Amount	1	922.68400417	16.79	0.0007
Material*Amount	2	26.42773333	0.24	0.7888

Figure 8 contains the results of this analysis. Note that for the contrasts we constructed the degrees-of-freedom, sums of squares, F statistics, and p-values are identical to those obtained by the standard SAS analysis. Note also that we used the SOLUTION option and the NOINT option on the model statement in our SAS program. The NOINT option is what forced the parameters of the model to become the cell means, and the SOLUTION option printed out these parameter estimates so that we can verify that they are indeed the cell means.

We can also perform a one-way CRD analysis on the separate SAS data set that contains the 12 observations from the three control groups, which differ in temperature, to see if the response variables differ. Figure 9 contains the computer program to perform a one-way CRD analysis of these 12 observations.

Figure 9. SAS program to do a one-way, 3-group ANOVA.

```
proc glm data=control;
  class group;
  model crack shatter = group;
  means group;
  manova h=group;
  title 'One-way CRD in the control groups';
run;
```

Note that we are using a different SAS data set which contains only the twelve observations in the three control groups, treating them as completely separate from the remaining observations and analyzing them in a completely independent analysis. One problem,

of course, is that we are unable to test any differences between the control groups and the groups that got tempering material since we are handling them in completely separate analyses. Figure 10 contains the output from this analysis for the dependent variable CRACK. You will notice that there appears to be no significant difference among these groups.

Figure 10. Univariate output for the variable CRACK.

Source	DF	Type III SS	F Value	Pr > F
GROUP	2	219.10531667	1.5	0.2713

We can perform the same analysis by fitting these 12 observations as a one-way, 3-group cell means model, CRD, and testing the difference among the 3 parameters with a two df contrast (Figure 11). (No output is shown.)

Figure 11. SAS program to fit a one-way 3-group design using a cell means model and test effects using a contrast statement.

```
proc glm data=control;
  class group;
  model crack shatter = group
    /solution noint;
  contrast 'One-Way CRD' group 1 -1 0,
    group 1 0 -1;
  means group;
  manova h=group;
  title 'One-way CRD in the control groups';
run;
```

Note that in this analysis, we are fitting a model with three cell means, one for each of three cells. To test if the three cell means are equal, we use a 2 df contrast matrix. The first row (1 -1 0) tests the equality of the first and second means, and the second (0 1 -1) tests the equality of the second and third means. The two rows are linearly independent, and if both differences equal zero, the three means are equal.

Optimal Analysis

A more optimal way of analyzing these data (Figure 12) uses the full data set, containing all nine groups and all thirty-six observations, and fits a nine group nine parameter cell means model to these data.

We then test hypotheses equivalent to the hypotheses tested in the previous two analyses using these nine groups. An hypothesis that would involve either tempering material amount or the interaction would involve only the first six groups. Since we have nine parameters or nine cell means any contrasts would themselves have nine coefficients. We would simply allow the three coefficients corresponding to the three groups not involving tempering material to be zero.

Thus, the first contrast, labelled Material, is a two df contrast exactly like the Material contrast in the two-by-three factorial we analyzed earlier, except that it has 9 coefficients instead of 6, because there are 9 cells in this design. Since this contrast does not involve the last three cells, the coefficients for those three cells are all zero. Thus, the contrast has 9 coefficients, the last three of which are zeros (.5 .5 -.5 -.5 0 0 0 0 0 for the first row, and 0 0 .5 .5 -.5 -.5 0 0 0 for the second row).

For the Amount effect, the 9 coefficients consist of the 6 which we saw in the two-by-three factorial used for the Amount effect, and three zeros so that the last three cell means do not enter into this contrast. The interaction, as before, was constructed using a horizontal direct product.

Any hypotheses just involving temperature, or really only the three groups in the control part of our design, would simply have zero coefficients for the six parameters corresponding to the cells not of interest and would use the same parameters we used when we analyzed the control data separately. Thus, the Temp contrast is a 2 df contrast that tests the equality among three parameters: the last three cell means. This two-row contrast matrix has the first 6 coefficients in each row all zero, so that these contrasts really only involve the last three cells, the ones that vary along temperature. Notice that both of these sets of hypotheses are in fact more efficiently tested using all nine cells because we have more subjects, therefore more degrees-of- freedom, and further a more precise estimate of the pooled within cells variance (means square error).

Figure 12. SAS program to fit a one-way 9-cell cell means model and test main effects, interactions, temperature effect, and tempered vs nontempered cells using contrasts.

```

proc glm data=full;
  class group;
  model crack shatter = group
    /solution noint;
  means group;
  contrast 'Material'
    group .5 .5 -.5 -.5 0 0 0 0 0,
    group 0 0 .5 .5 -.5 -.5 0 0 0;
  contrast 'Amount'
    group .33333 -.33333 .33333 -.33333
      .33333 -.33333 0 0 0;
  contrast 'Material*Amount'
    group .5 -.5 -.5 .5 0 0 0 0 0,
    group 0 0 .5 -.5 -.5 .5 0 0 0;
  contrast 'Temp'
    group 0 0 0 0 0 0 1 -1 0,
    group 0 0 0 0 0 0 0 1 -1;
  contrast 'anything vs cntl'
    group .16667 .16667 .16667 .16667 .16667 .16667
      .16667 -.33333 -.33333 -.33333;
  contrast 'anything vs 1000'
    group .16667 .16667 .16667 .16667 .16667 .16667
      .16667 0 -1 0;
  manova h=group/short;
  title 'Full Analysis'
run;

```

Finally, we are able to do contrasts in this analysis that we were not able to do in either of the other analyses. For example, we can construct a contrast testing the difference between the mean of the groups that received tempering material of any sort in any amount (the six groups from the factorial part of the design) and the group that received no tempering material and was fired at 1000 degrees. The contrast labelled "Anything vs 1000" tests that hypothesis. The first six coefficients (.16667 .16667 .16667 .16667 .16667 .16667) take the mean of the first 6 cell means, and the last three coefficients (0 -1 0) pick up the mean for the cell which received no tempering material and was fired at 1000 degrees.

Similarly, we can construct a contrast testing the difference between the mean of the groups that received tempering material of any sort in any amount (the six groups from the factorial part of the design) and the mean of the three groups that received no tempering

material. The contrast labelled "Anything vs ctrl" tests that hypothesis. The first six coefficients (.16667 .16667 .16667 .16667 .16667 .16667) take the mean of the first 6 cell means, and the last three coefficients (-.33333 -.33333 -.33333) take the mean of the last three cell means, the cells which received no tempering material.