Interaction Effects: Their Nature and Some Post Hoc Exploration Strategies

James M. Graham

Texas A & M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 28, 2000.

Abstract

The present paper reviews the basics of the elusive but important concept of the interaction effect. The ANOVA interaction effect is defined both conceptually and mathematically. Common errors in interpreting interaction effects are discussed and appropriate strategies for achieving post hoc understandings of the origin of detected interaction effects are presented.

In his 1957 American Psychological Association presidential address, Lee J. Cronbach argued the importance of investigating whether a given psychological or educational intervention works best for everyone, and, if not, what interventions work best for which types of people. Cronbach referred to these studies as aptitude-treatment interaction (or ATI) studies.

Since then, ATI studies have been utilized with increasing frequency (Dodds, 1998). Of special interest to the researcher employing ATI designs is the analysis of variance (or ANOVA). This statistical method, first introduced by R. A. Fisher in 1915, allows the simultaneous examination of not only multiple independent variables but also of the interactions between those independent variables (Marascuilo & Serlin, 1988).

Interaction effects are of special interest to researchers, as few presume that any given educational or psychological intervention will work equally well for every individual. The frequency of statistically significant interactions and the importance ascribed to those interactions in the research community show the need for a more detailed understanding of the ANOVA interaction effect (Rosenthal & Rosnow, 1984). Many researchers regard the ANOVA as simply a calculation performed by a computer, carried out mechanically with little to no understanding of what it entails. However, in the case of the ANOVA, a full understanding is necessary in order to successfully apply the ANOVA technique, particularly in regard to interaction effects (Huitson, 1971). The decided lack of understanding demonstrated by many investigators is responsible for the preponderance of errors in interpreting interaction effects found in current literature.

Indeed, Rosnow and Rosenthal (1989b) have referred to the results of interaction effects as “probably the universally most misinterpreted empirical results in psychology” (p. 1282). In fact, in a review of 191 research articles using ANOVA designs in prominent journals Rosnow and Rosenthal (1989) found that only 1% of articles correctly interpreted interaction effects!

Throughout this paper 2x2 ANOVA designs are used as examples. It is important to note that the principles presented here apply to all ANOVA designs; 2x2 designs are used as examples only because they are convenient and simple.

Definition of Interaction Effects

While there are nearly as many definitions of the interaction effect as there are textbooks and theoretical articles, all describe the same basic characteristics. For the purposes of this paper, Keppel's (1991) description of interactions will suffice: "An interaction is present when the effects of one independent variable on behavior change at the different levels of the second independent variable" (p. 196). When the differences in the dependent variable between experiment conditions are not accounted for by only the main effects of the independent variables and by error, those differences not accounted for are said to occur due to an interaction between the two independent variables. Interactions are therefore sometimes referred to as "residual" effects, or the effects remaining after lower-order effects (the main effects of the independent variables) are removed (Rosenthal & Rosnow, 1984; Rosnow & Rosenthal, 1991). A greater understanding of this concept can be gained by examining the mathematical model on which the two-way ANOVA is based.

ANOVA Mathematical Model

The model on which the two-way ANOVA is based posits that each observation or condition can be partitioned into five distinct components (Marascuilo & Levin, 1970). In a balanced design each of these components is wholly uncorrelated and unrelated to each of the others (Hester, 1996). Using the notation of Marascuilo and Levin (1970), these components can be shown as:

$y_{ijk} = (\mu + \alpha_i + \beta_j + \gamma_{ij}) + e_{ijk}$,

where: μ is the grand mean of all observations; $\alpha_i = \mu_i - \mu$ = the difference between the mean of all cases tested under the same A condition as this group and the grand mean; $\beta_j = \mu_j - \mu$ = the difference between the mean of all cases tested under the same B condition as this group and the grand mean; $\gamma_{ij} = \mu_{ik} - \mu_i - \mu_j + \mu$ = the mean for all cases in condition A_iB_j, less the mean of all cases tested under the same A condition as this group, less the mean of all cases tested under the same B condition as this group, plus the grand mean; and e_{ijk} = error.

An observation or condition in a two-way ANOVA is therefore made up of the effects of the grand mean, independent variable A, independent variable B, the interaction between independent variables A and B, and error. Again, each of these partitions (with the exception of the grand mean) are completely uncorrelated with one another in a balanced design. Knowing the effects of the main independent variables tells you nothing about the effects of the interaction (Hester, 1996). This concept, while difficult for many to grasp, is key to the understanding of interaction effects.

Interaction Hypothesis in ANOVA

When testing for the presence of a statistically significant interaction effect, a 2-way ANOVA tests the hypothesis that the interaction effects (γ_{ij}) for each experiment condition are equal to zero (Levin & Marascuilo, 1972). Put another way, H_0: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{ij} = 0$. Using the mathematical model underlying the ANOVA after removing error, the interaction effect for a given condition can be defined as $\gamma_{ij} = y_{ijk} - \mu - \alpha_i - \beta_j$. When the γ for each experimental condition is found to be equal to zero, the null hypothesis of zero interaction is not rejected. When the γ for one or more of the experimental conditions is found to not equal zero, the null hypothesis is rejected and a statistically significant interaction is said to exist.

It is of the utmost importance for the researcher to understand that when interaction is tested, the ANOVA is testing the hypothesis that the interaction effects for each experimental condition are statistically equal to zero. Many common errors made when interpreting interaction effects are based on a misunderstanding or neglect of this principle.

Example

In order to better understand what interaction effects are, consider the table of population means for a 2x2 ANOVA design presented in Table 1. In this table, the means for each combination of independent variables is shown, as well as the means for each row and column and the grand mean. The numbers in parentheses are the main effects (α_i and β_j), or the differences between the row or column means and the grand mean ($\alpha_1 = \mu_{1j'} - \mu$; $\alpha_2 = \mu_{2j'} - \mu$; $\beta_1 = \mu_{i'1} - \mu$; $\beta_2 = \mu_{i'2} - \mu$). This set of means represents a population in which there is no interaction between factor A and factor B. A lack of interaction is often signified by parallel lines in a plot of cell means (Hinkle, Wiersma, & Jurs, 1998), as shown by the parallel lines in Figure 1. Note that there is no interaction if the lines are parallel even if the lines are not horizontal. Each condition, or cell, mean (μ_{ij}) can be derived by taking the grand mean and adding the appropriate row (α_i) and column (β_j) effects. For example, consider the population mean defined by row one and column one. When applied to the equation $\mu_{ij} = \mu + \alpha_i + \beta_j$, we see that it is a perfect fit: 7 = 5 + (+1) + (+1), 7 = 7. This is the case for each of the cell means.

INSERT TABLE 1 AND FIGURE 1 ABOUT HERE

Now consider the table of population means presented in Table 2. When applying the formula for determining cell means to the population defined by row one column one (μ11), we see that: 2 does not equal 4.5 + (-1.5) + (+.5), or 2 does not equal 3.5. In this condition there is an effect of -1.5 that is not accounted for by the grand mean or either of the main effects. This “leftover” or residual effect is what is commonly called the interaction effect.

INSERT TABLE 2 ABOUT HERE

In the case of the population defined by row one, column one, the interaction effect of –1.5 is due, not to either of the independent variables alone, but to an interaction between the two independent variables. As our original formula no longer fits when an interaction is present, we must adjust the formula by adding the interaction effect (γ_{ij}): $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$. A plot of cell means for this new example, as shown in Figure 2, shows that the lines are not parallel, thus signifying the presence of an interaction. Solving our new equation for the interaction effect gives us the same mathematical definition of the interaction effect given by the mathematical model of the ANOVA: $\gamma_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$.

INSERT FIGURE 2 ABOUT HERE

Interpreting Interaction Effects

Type IV Errors

Researchers have long had to contend with the possibility of committing Type I and Type II errors while conducting research. Of special interest to the researcher working with interaction effects, however, is what have been termed Type IV errors. A Type IV error is said to occur when a hypothesis is correctly rejected but incorrectly interpreted (Levin & Marascuilo, 1972; Marascuilo & Levin, 1970). As applied to interaction effects, a Type IV error is committed when post hoc examinations of correctly identified statistically significant interactions neither test the original hypothesis nor fit with the underlying model upon which the ANOVA is based (Dodds, 1998; Levin & Marascuilo, 1972).

Incorrect Use of Post Hoc Tests

When an ANOVA design is used and statistical significance for any effect is found, post hoc tests are applied to determine the location of the differences between experimental conditions. Post hoc tests are not applied, however, in the case of 2x2 ANOVA designs, as the statistically significant effects can only be present in the one single possible contrast (Dodds, 1998). The post hoc tests developed by Tukey and Sheffe' are often used to examine interaction effects. These tests are used to make comparisons of condition (or cell) means in an attempt to locate the source of an effect.

The Tukey method is designed to only make simple comparisons of means (Hinkle et al., 1998). For example, when used to interpret ANOVA results for the A main effect in a 3X3 design, Tukey's method tests the following hypothesis for each pairwise comparison:

$H_0 : \mu_i = \mu_k$ for i not equal to k, or

$\mu_{1j} = \mu_{2j}$; $\mu_{2j} = \mu_{3j}$; $\mu_{3j} = \mu_{1j}$

Likewise, while examining the B main effect in the same design, Tukey's method tests all possible combinations of the second independent variable, or: $\mu_{i1} = \mu_{i2}$; $\mu_{i2} = \mu_{i3}$; $\mu_{i3} = \mu_{i1}$ (Keppel, 1991).

All contrasts/comparisons always test differences between only two means. However, "simple" contrasts compare the means of a given two levels of a way *without combining* any levels. For example, in a 4x2 design (4 undergraduate class levels [freshman, sophomore, junior, senior] and 2 gender groups [male, female]) with 10 people in each of the eight design cells, the comparison of the dependent variable mean of the 20 freshman with the dependent variable mean of the 20 sophomores is a “simple” contrast.

Tukey’s method, as it is commonly (though erroneously) applied to interpreting interaction effects, tests all possible combinations of both independent variables, or $\mu_{11} = \mu_{12}$; $\mu_{12} = \mu_{21}$; …; $\mu_{ij} = \mu_{kk}$ (Harwell, 1998). In the same manner, Scheffe’s method can be used, not only to test the simple comparisons, but the “complex” contrasts as well (Keppel, 1991).

“Complex” contrasts include “simple” contrasts , but also create contrasts of two means in which one or both means are computed by aggregating data from more than one level. For example, in the previous example of a 4x2 design, the contrast of the dependent variable mean of the 20 freshman with the dependent variable mean of the 40 juniors and seniors combined is a “complex” contrast. The contrast of the dependent variable mean of the 40 freshman and sophomores combined with the dependent variable mean of the 40 juniors and seniors combined also is a “complex” contrast.

It is the erroneous application of post hoc tests that account for the preponderance of misinterpretations of interaction effects. The use of post hoc tests to examine conditional or cell means is not based on the mathematical formula underlying the ANOVA model, and therefore may be inaccurate. In addition, the use of cell means in post hoc tests does not test interactions and does not even test the original ANOVA hypothesis.

Testing condition means does not test interactions. The practice of comparing conditional or cell means to interpret interaction effects is a grievous error. In doing so, the researcher is looking at pairs of cell means to determine where the differences in the dependent variable lie in an attempt to locate the source of interaction. These cell means, while they contain the effects of the interaction, also contain a host of other effects which prevent accurate interpretation of interactions. As stated by Rosnow and Rosenthal (1989a), these cell means “are the combined effects of the interaction, the row effects, the column effects, and the grand mean” (p. 144).

To make this more clear, consider the mathematical explanation of the cell mean, $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$. This shows that the cell means contain effects caused not only by the interaction, but are also influenced by the grand mean and the main effects. To say that a comparison of cell means tests pairs of interactions is simply not true. A comparison of cell means tests pairs of interactions, plus a host of other confounding information. In order to correctly interpret interaction effects, it is necessary to employ post hoc tests to examine pairs of interaction effects, not pairs of cell means.

The use of simple effects, or cell means, to interpret interaction effects can be very misleading. The main effects of the independent variable may contribute to the cell means even more than the interactions (Rosnow & Rosenthal, 1989a). This error is greatest when the lower order (main) effects are statistically significant (Harwell, 1998). In this case, the use of simple effects to interpret interactions may actually be more accurate in interpreting the main effects. As all effects in a balanced ANOVA design are uncorrelated, this would reveal absolutely nothing about the interaction effects which are the original focus of the investigator.

Testing cell means does not test the original hypothesis. In addition to the practice of using cell means to interpret interactions not actually testing interactions, the hypotheses tested by simple effects tests are not consistent with the original hypothesis tested by the ANOVA (Boik, 1979; Levin & Marascuilo, 1972). The original interaction hypothesis, as specified by the ANOVA, is $H_0 : \gamma_{11} = \gamma_{12} = \ldots = \gamma_{ij} = 0$. Using cell means to interpret interactions does not test this hypothesis, rather it tests the hypothesis $H_1 : \mu_{11} = \mu_{12} = \ldots = \mu_{ij}$.

With the knowledge that γ_{ij} does not equal μ_{ij}, it becomes clear that the use of cell means to interpret interactions does not only not test interactions but also has nothing to do with the original hypothesis. When a statement about interactions based on post hoc comparisons of cell means is made, a Type IV error has been committed. In this case, the correctly rejected hypothesis has been interpreted in a manner inconsistent with the original ANOVA interaction hypothesis.

Correct Use of Post Hoc Tests

When a researcher attempts to interpret statistically significant interaction effects, it is important that the post hoc tests employed test the interactions, not the cell means (Rosnow & Rosenthal, 1991). This is most often accomplished by using simple effects tests, not of the original cell means, but of “corrected” cell means that more accurately reflect the interactions (Rosnow &

Rosenthal, 1989a). While it is not necessary to correct cell means if the main effects are zero, the cell means must be adjusted if even one of the main effects is found to be greater than zero (Harwell, 1998). As it does not harm one's interpretation to correct for main effects that are zero, it is good practice to adjust cell means regardless of the statistical significance of main effects.

Computing corrected cell means. The process for obtaining interaction effects for use in post hoc tests is simply a solution of the basic ANOVA model for the interaction effects (Meyer, 1991). Rosnow and Rosenthal (1989a) have provided the following simple formula for determining the corrected cell means:

$$\text{Adjusted } \mu_{ij} = \mu_{ij} - \alpha_i - \beta_j - \mu,$$

where μ_{ij} = the original cell mean; $\alpha_i = (\mu_i - \mu)$ = the effect of level i of factor A = mean of all cases in level i less the grand mean; $\beta_j = (\mu_j - \mu)$ = the effect of level j of factor B = mean of all cases in level j less the grand mean; and μ = the grand mean. This formula for producing corrected cell means, expressed a different way, has been simplified by Harwell (1998) for ease of use:

$$(\mu_{ij} - \mu) - (\mu_i - \mu) - (\mu_j - \mu) =$$

$$\mu_{ij} - \mu_i - \mu_j + \mu$$

For example, consider the plot of fictional population means for a 2x2 design presented in Table 3.

INSERT TABLE 3 ABOUT HERE

In Table 3, $\mu 11$ = 5 is the mean of the population defined by the row one, column one cell; $\mu 21$ is the mean of the population defined by the row two, column one cell; etc. Population means for each of the rows and columns are also presented, as well as the grand mean. Using the computational formula presented by Harwell (1998), it is a simple matter to determine the corrected cell means shown in Table 4:

Corrected $\mu_{11} = 5 - 6 - 7 + 6 = -2$

Corrected $\mu_{12} = 7 - 6 - 5 + 6 = 2$

Corrected $\mu_{21} = 9 - 6 - 7 + 6 = 2$

Corrected $\mu_{22} = 3 - 6 - 5 + 6 = -2$

INSERT TABLE 4 ABOUT HERE

While this example involved a 2x2 design, the same process can be used to determine the adjusted cell means for any 2-way ANOVA design. These corrected cell means, while easy to compute, are often neglected by researchers. The corrected means provide the key which allows interaction effects to be accurately interpreted. It may be of special interest to the student of interaction effects that, as Harwell (1998) pointed out, these adjusted cell means "when squared, weighted by the number of scores in a cell, and summed produce the sum of squares interaction SS(AB) used to compute the mean square interaction term for the numerator of the F test" (p. 128).

Using corrected means in post hoc tests. All post hoc comparisons of interaction effects must be interaction contrasts, and not contrasts of cell means (Keppel, 1991; Marascuilo & Levin, 1970). Performing interaction contrasts can be done easily by performing the same simple effects tests used to interpret main effects. However, it is of the utmost importance that these tests be used to contrast, not condition means (μ_{11}, μ_{12}, μ_{21}, …, μ_{ij}), but condition interactions (γ_{11}, γ_{12}, γ_{21}, …, γ_{ij}). This ensures that the post hoc test hypotheses are consistent with the original ANOVA interaction hypothesis and are accurately interpreting only the interactions.

Adjusted cell means can be used to test both simple interaction effects and complex interaction effects in the same way that row and column means can be used to test simple and complex main effects. Instead of testing all possible combinations of means it is appropriate to test all possible combinations of interactions, to eliminate the possibility of the main effects confounding the interpretation of the interaction. Marascuilo and Levin (1970) provided a detailed description of how to use Scheffe's method to test interpret interactions, with a number of worked examples.

The principle, though often ignored, is a simple one. The use of pairwise comparisons of condition means to test interactions is a Type IV error and should be avoided. Proper methodology dictates that post hoc tests make use of pairwise comparisons of interaction effects, or, as we have seen, corrected cell means.

Plotting Interactions

Another common method of interpreting ANOVA results is the graphical plotting of cell means. This is considered good practice, as the cell means are often of great interest to researchers. The majority of researchers, however, incorrectly refer to these plots of cell means as plots of interactions and use these plots to interpret the interaction (Rosnow & Rosenthal, 1991). As previously discussed, cell means contain information not only about the interaction effect, but about the main effects as well.

To attempt to interpret the interaction effects by using plots of cell means is to make the same Type IV error committed by using simple effects tests of cell means to interpret interactions, unless the main effects are zero.

To make this discussion more clear, consider the plot of cell means presented in Figure 2. In discussing the interaction a researcher may infer that the performance of participants in condition a_1 increases slightly across levels of B, while those in condition a_2 decrease severely. This statement, however, is not an interpretation of only the interaction effect, but of the interaction effect <u>and</u> the main effects.

In order to accurately represent interaction effects graphically, investigators must plot, not the cell means, but the <u>corrected</u> cell means, or the interaction effects themselves (Harwell, 1998; Rosnow & Rosenthal, 1989a). The corrected cell means can be easily calculated as previously discussed, and plotted in the same way as cell means. The corrected cell means for this example have been plotted in Figure 3. An examination of this figure reveals that, contrary to the previous interpretation, the performance of participants in condition a_1 increases across levels of B to the exact same degree that the performance of participants in condition a_2 decreases across levels of B. The "slight" and “severe” changes across levels of B pointed out in Figure 2 are artifacts of the main effects. In actuality, the interaction shows that participants in condition a_1 are helped by treatment b_2 to the exact same extent that participants in condition a_2 are hurt by it.

INSERT FIGURE 3 ABOUT HERE

Discussion

Many factors may contribute to the widespread misunderstanding and misinterpretation of interaction effects. The fact that many textbooks either avoid discussing the interpretation of interactions or erroneously state that they can be interpreted by examining only cell means is likely a major contributor (Harwell, 1998; Rosnow & Rosenthal, 1991). While corrected means are relatively easy to calculate, many researchers may not do so simply because the vast majority of data-analytic software packages (SPSS, SAS, BMDP, MINITAB) neither compute nor plot adjusted cell means (Harwell, 1998; Rosnow & Rosenthal, 1989b). Certainly the editorial policies of many journals, which permit the publication of misinterpreted data, are also to blame (Harwell, 1998).

Regardless, it is of the utmost importance that researchers in fields of education and psychology learn to correctly report and interpret interaction effects when using an ANOVA design. The practice of using unadjusted cell means to do so is erroneous, as it results in an inaccurate representation of interactions. While Type I and Type II errors in some circumstances may be unavoidable, the same is not true of Type IV errors. The incorrect interpretation of a correctly rejected hypothesis is preventable with a thorough understanding of interaction effects.

References

Boik, R. J. (1979). Interactions, partial interactions, and interaction contrasts in the analysis of variance. Psychological Bulletin, 86, 1084-1089.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Dodds, J. (1998, November). Understanding interaction effects and Type IV errors. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 426 094)

Harwell, M. (1998). Misinterpreting interaction effects in analysis of variance. Measurement and Evaluation in Counseling and Development, 31, 125-136.

Hester, Y. (1996, January). Understanding that ANOVA effects are perfectly uncorrelated. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 397 088)

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). Applied statistics for the behavioral sciences (4th ed.). Boston: Houghton Mifflin Company.

Huitson, A. (1971). The analysis of variance: A basic course. London: Griffin.

Keppel, G. (1991). Design and analysis: A researcher's handbook (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Levin, J. R., & Marascuilo, L. A. (1972). Type IV errors and interactions. Psychological Bulletin, 78, 368-374.

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. American Educational Research Journal, 7, 397-421.

Marascuilo, L. A., & Serlin, R. C. (1988). Statistical methods for the social and behavioral sciences. New York: W. H. Freeman and Company.

Meyer, D. L. (1991). Misinterpretation of interaction effects: A reply to Rosnow and Rosenthal. Psychological Bulletin, 110, 571-573.

Rosenthal, R., & Rosnow, R. L. (1984). Essentials of behavioral research: Methods and data analysis. New York: McGraw-Hill.

Rosnow, R. L., & Rosenthal, R. (1989a). Definition and interpretation of interaction effects. Psychological Bulletin, 105, 143-146.

Rosnow, R. L., & Rosenthal, R. (1989b). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at only the interaction (unless all main effects are zero). Psychological Bulletin, 110, 574-576.

Table 1

Cell Means and Main Effects for a 2x2 ANOVA Design With No Interaction

Level/ Effect	Level b_1	b_2	μ_i (α_j)
a_1	7	5	6 (+1)
a_2	5	3	4 (-1)
μ_j	6	4	μ_T = 5
(β_i)	(+1)	(-1)	

Table 2

Cell Means and Main Effects for a 2x2 ANOVA Design With Interaction

Level/ Effect	Level b_1	b_2	μ_i	(α_j)
a_1	2	4	3	(-1.5)
a_2	8	4	6	(+1.5)
μ_j	5	4	$\mu_T = 4.5$	
(β_i)	(+.5)	(-.5)		

Table 3

Uncorrected Cell Means for a 2x2 Factorial Design

Level/ Effect	Level b_1	b_2	μ_i
a_1	5	7	6
a_2	9	3	6
μ_j	7	5	μ_T = 6

Table 4

Cell Means for a 2x2 Factorial Design Corrected for Both Main Effects and the Grand Mean

Level	Level b_1	b_2
a_1	-2	2
a_2	2	-2

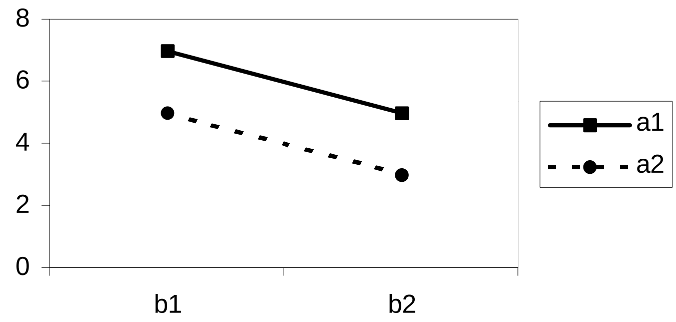


Figure 1. Plot of cell means for a 2x2 ANOVA design with no interaction.

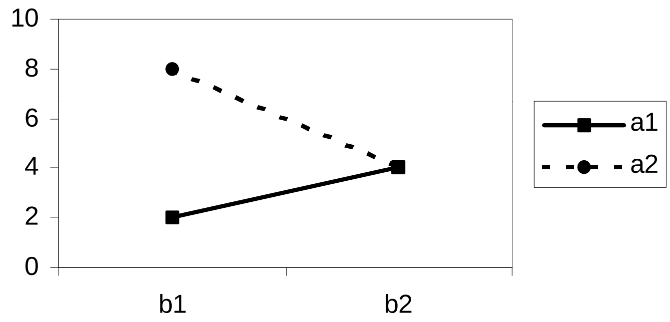
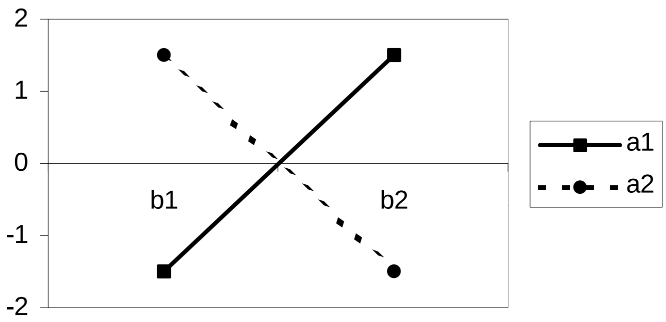


Figure 2. Plot of cell means for a 2x2 ANOVA design with interaction.



<u>Figure 3.</u> Interaction plot of corrected cell means.