DOCUMENT RESUME

ED 049 277 TM 000 449

AUTHOR Levin, Joel R.; Marascuilo, Leonard A.
TITLE Interactions Revisited.
PUB DATE Feb 71
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Analysis of Variance, Behavioral Science Research, Evaluation Techniques, *Hypothesis Testing, *Interaction, Mathematical Models, Orthogonal Rotation, Predictive Measurement, *Research Methodology, *Statistical Analysis

ABSTRACT

Marascuilo and Levin's (1970) notion of Type IV errors is extended, with respect to the interpretation of interactions in analysis of variance (ANOVA) designs. To help clarity what an interaction is and what it is not, in terms of the ANOVA model, the following points are made: (i) interactions should be thought of as linear contrasts involving particular cell means; (ii) such contrasts may be both specified and directional, in that they may be defined to test an investigator's a priori hypotheses; and (iii) even the layman's conceptualization of interactions fits nicely into the ANOVA model. (Author)

INTERACTIONS REVISITED[1,2]

Joel R. Levin
University of Wisconsin

and

Leonard A. Marascuilo
University of California, Berkeley

Paper presented at the annual meeting of The American Educational Research Association, New York, 1971.

Interactions Revisited[1,2]

The rejection or denial of a true statistical hypothesis and the nonrejection or acceptance of a false hypotheisis are abstractions with which researchers in the behavioral sciences generally make only passing acquaintance. While one is taught to express concern for -- and even to compute probabilities associated with -- Type I and Type II errors as a student attending a first course in statistics, the same student in the "cruel world" of research experiences understandable difficulty in deciding when either of these errors has in fact occurred. As a result, one gradually learns to live with them. Such should not be the case with Type IV errors, as introduced by Marascuilo and Levin (1970), since with practice this kind of error is easily recognized and consequently avoided.

According to the definition of Marascuilo and Levin, a Type IV error is said to occur whenever a correct statistical test has been performed, but is then followed by analyses and explanations which are not related to the statistical test used to decide whether the hypothesis should or should not have been rejected. More succinctly, a Type IV error is made whenever a researcher offers an incorrect interpretation to a correctly rejected statistical hypothesis. Less succinctly, a Type IV error is identified as having been committed whenever a researcher concludes, on the basis of an appropriately performed statistical test, that there is a reliable source of variability in the data, but then proceeds to

specify the locus of the effect with an eyeball interpretation of the data or by employing post hoc multiple comparison procedures which are not congruent with the hypothesis initially tested, and which may not even correspond to the underlying model upon which the statistical test was based.

Type IV Errors in the One-Way Analysis of Variance Model

Perhaps the most commonly encountered Type IV error is the one committed by a researcher who follows a rejected anlaysis of variance (ANOVA) hypothesis with a set of overlapping multiple *t*-tests, each performed at the same alpha level as chosen for the original *F*-test. In this case, the statistical hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_I$ has been rejected with the probability of a Type I error set equal to α. If the researcher now examines each of the $Q = \frac{I(I-1)}{2}$ paired mean comparisons at the same α level as used for the *F*-test, the total probability of at least one Type I error in the *set* of comparisons is inflated far above the original probability to a maximum value of $\alpha_0 \leq Q\alpha$. Should this procedure pronounce statistically significant certain pairwise comparisons that would not have been identified with the "appropriate" post hoc Scheffe' (1953) method, then a Type IV error would have been made.[3]

Some researchers may attempt to control the maximum probability of at least one Type I error in the set of Q pairwise contrasts (or in the set of any K planned comparisons, i.e., contrasts involving linear combinations of means, as well as pairwise comparisons) by

using Bonferroni critical values as described by Miller (1966) or Dunn (1961) to conduct their post hoc investigations. However, even these adjustments do not eliminate Type IV errors since they are not related to the original F-test in a one-to-one manner. Because the Bonferroni procedures are quite powerful -- especially when the number of comparisons of interest is relatively small -- they may detect a greater number of significant differences than would the Scheffe´ method, which corresponds exactly to the classical F-test.

It is worth mentioning that Type IV errors of this kind may be avoided simply by bypassing the F-test altogether. In point of fact, abandoning the F-test might be the optimal strategy for a researcher to follow if the plan is to examine only a small number of contrasts, since the F-test could lead to a nonrejection of H_0 while one or more comparisons could be identified as significant with the more powerful Bonferroni or Dunn method. In such cases, the relative power of the Scheffe´ to the Dunn procedure (defined, perhaps, in terms of the ratio of the respective critical values) may be determined prior to data collection, in order to reach a rational decision concerning the approach to adopt (see Davis, 1969).

It is also worth noting that in the equal sample size model, employing the Tukey (1953) method of pairwise comparisons following the rejection of H_0 based on the classical F-test may also produce Type IV errors, since Tukey's procedure is more likely to identify pairwise differences as significant than is Scheffe´'s. The reason

for this potential discrepancy is that the two procedures are derived from different underlying distributions. Tukey's more powerful method is based on the distribution of the studentized range for which the corresponding test statistic is the ratio of the weighted (by the square root of the common sample size) maximum mean difference to the square root of the mean square within. Scheffe´'s procedure, on the other hand, is based on a different mathematical model for which the test statistic is the familiar ratio of the mean square between to the mean square within in the one-way ANOVA model. Thus, if a researcher is interested in performing only pairwise contrasts, and if sample sizes are equal, then the F-test is not the "appropriate" test to perform! In this case, the researcher should first perform the studentized range test and if it leads to rejection of H_0, then Tukey's method of pairwise comparisons should be employed (Dixon and Massey, 1969; Scheffe´, 1959). With this strategy, the probability of making a Type IV error is reduced to zero.

Type IV Errors in Other Models

Cautions regarding the incorrect post hoc analysis of correctly performed omnibus tests are not confined to traditional ANOVA designs. Marascuilo (1966) has described simultaneous inference procedures which are "appropriate" for large-sample tests of the differences among J independent proportions and among J independent correlation coefficients. Steel (1961), Dunn (1964), and Marascuilo and McSweeney (1967) have developed multiple comparison techniques

to accompany the nonparametric rank tests. The essential feature of such post hoc methods is that they are <u>based on the same distribution as the test statistic</u>, and therefore will yield information which is congruent with the test initially performed.

As might be surmised, multivariate ANOVA hypotheses offer the researcher a Pandora's box of very elegant and sophisticated Type IV errors (among others) that may result in unwarranted explanations of significant findings. One common error following a rejected multivariate hypothesis is to perform variable by variable comparisons (perhaps at a reduced α level, or using post hoc univariate techniques), or to interpret the significant multivariate statistic in terms of linear combinations of dependent variables, as might be suggested from an examination of principal components or linear discriminant functions. While there may be some correspondence between the decisions made under these analysis procedures and the "appropriate" Roy-Bose multivariate post hoc method as described by Morrison (1967), it has been shown by Hummel and Sligo (in press) that the Type I error probabilities of the multivariate and univariate procedures are not identical. That is, when multivariate data are analyzed on a post hoc basis with critical values determined from univariate procedures, one is liable to arrive at statistical decisions which are different from those based on multivariate post hoc techniques. As was mentioned for the univariate case, if only a small number of comparisons is of interest, it might be advisable to examine those comparisons

individually rather than to perform an overall multivariate F-test.

Until now, the discussion of Type IV errors has been basically from a Type I error point of view. In other words, situations were described in which the correspondence between the Type I error probability of the post hoc analyses and that of the test initially performed was less than perfect. However, another (perhaps more serious) Type IV error occurs when a researcher defines his post hoc contrasts in a manner which does not even investigate the hypothesis he is presumably testing. This kind of error frequently occurs when it comes to interpreting statistical "interactions" in contingency tables (Goodman, 1964; Marascuilo, 1966), regression analyses (Timm, in preparation), and factorial ANOVA designs (Marascuilo and Levin, 1970).

Marascuilo and Levin point out that a typical strategy following the detection of a significant interaction in a factorial ANOVA is to make either pairwise or nested comparisons using the various cell means of the design. They showed by examples that this procedure is in no way related to the interaction F-test initially performed. It was further suggested that this error arises because many researchers do not have a clear understanding of what constitutes an interaction as it is defined by the mathematical ANOVA model. As a result, the post hoc procedures and/or verbal discussion based on the identification of a significant interaction often are inappropriate. A further clarification of the meaning of a statistical interaction will be attempted in the sections which follow.

Mathematical Model for an I by J Factorial Design

Consider a two-way fixed effects ANOVA model with an equal number of observations per cell.[4] For I rows and J columns, the model may be written as:

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

where:

y_{ijk} = the value of the k-th observation in the i-th row and j-th column

μ = a fixed constant that centers the data

α_i = the effect of Level i for Factor A where $\sum_{i=1}^{I} \alpha_i = 0$

β_j = the effect of Level j for Factor B where $\sum_{j=1}^{J} \beta_j = 0$

γ_{ij} = the joint effect of Level i and Level j where $\sum_{i=1}^{I} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0$

e_{ijk} = the error associated with the k-th observation in the ij-th cell.

The e_{ijk} are assumed to be statistically independent, normally distributed, with a mean of zero and a variance equal to σ^2.

Under this model it is customary to test for the presence of row effects, column effects, and interaction effects by means of three orthogonal tests of hypothesis:

H_{01}: $\alpha_1 = \alpha_2 = \dots = \alpha_I = 0$

H_{02}: $\beta_1 = \beta_2 = \dots = \beta_J = 0$

H_{03}: $\gamma_{11} = \gamma_{12} = \dots = \gamma_{IJ} = 0$

If any of the main effects is signitifant, one may use Scheffe´'s method to identify possible sources of variance in exactly the same manner as that used for the one-way ANOVA. However, if the interaction test leads to a rejection of H_{03}, slightly different procedures must be used to locate the significant sources that account for the rejection of the hypothesis.

As suggested by Marascuilo and Levin (1970), the common practice of making simple comparisons among the cell means to locate the sources of a significant interaction is not valid if H_{03} is rejected. The procedure is appropriate only if instead of testing H_{01}, H_{02}, and H_{03} as orthogonal hypotheses, one were to test the composite total cell hypothesis:

H_{04}: $\mu_{11} = \mu_{12} = \dots = \mu_{IJ}$.

In this case, the Scheffe´ coefficient would be given by:

$$\underline{S} = \sqrt{(IJ-1)\underline{F}_{(IJ-1),IJ(n-1)}(1-\alpha)}$$

where n = number of observations per cell and α = the probability of a Type I error associated with H_{04}. In this case, a typical pairwise contrast is defined by:

$$\Psi = \mu_{ij} - \mu_{i'j'}$$

$$= (\mu + \alpha_i + \beta_j + \gamma_{ij}) - (\mu + \alpha_{i'} + \beta_{j'} + \gamma_{i'j'})$$

$$= (\alpha_i - \alpha_{i'}) + (\beta_j - \beta_{j'}) + (\gamma_{ij} - \gamma_{i'j'})$$

Note, however, that if this contrast were found to be significant, it would be impossible to know whether or not the difference was due to the fact that $\alpha_i \neq \alpha_{i'}$, or that $\beta_j \neq \beta_{j'}$, or that $\gamma_{ij} \neq \gamma_{i'j'}$, or any combination of these. In words, contrasts of this type lead to a confounding of the model's parameters, and the situation is not alleviated simply by testing H_{04} instead of H_{03}.

On the basis of this discussion it should not be concluded that there is no appropriate way to interpret the meaning of a significant F-ratio for interaction, in terms of some linear combination of cell means. What the previous discussion is meant to suggest is that linear contrasts of the form $\Psi = \mu_{ij} - \mu_{i'j'}$ that are typically defined by researchers to interpret interactions are incorrect in a Type IV error sense.

Interaction in the 2 by 2 Design

The simplest way to consider the problem and identify valid interaction contrasts is to reexamine a factorial design with I=2 and J=2. For this design, the observed grand, row, column, and cell means may be denoted as shown in Table 1.

Insert Table 1 about here

With four cells, there are three degrees of freedom available for the between cell hypothesis H_{04}. While the test of H_{04} could be performed, the more usual approach is to take the total sum of squares between groups and partition it into three orthogonal components each possessing one degree of freedom and each leading to an F-test with $\nu_1 = 1$ and $\nu_2 = 4(n-1)$. Although there is an unlimited number of ways that could be used to partition the sum of squares between groups, in factorial designs the partitioning consists of three very specific orthogonal contrasts. In this case, the two orthogonal contrasts for the A and B main effects are respectively given by:

$$\hat{\Psi}_A = \bar{y}_{1.} - \bar{y}_{2.} = (+1/2)\bar{y}_{11} + (+1/2)\bar{y}_{12} + (-1/2)\bar{y}_{21} + (-1/2)\bar{y}_{22}$$

$$\hat{\Psi}_B = \bar{y}_{.1} - \bar{y}_{.2} = (+1/2)\bar{y}_{11} + (-1/2)\bar{y}_{12} + (+1/2)\bar{y}_{21} + (-1/2)\bar{y}_{22}$$

To generate the third contrast orthogonal to both $\hat{\Psi}_A$ and $\hat{\Psi}_B$, one needs only multiply the coefficients pair by pair and use the resulting products as the coefficients for the third contrast. For the 2 by 2 design, this procedure defines the contrast:

$$\hat{\Psi}_{AB} = (+1/4)\bar{y}_{11} + (-1/4)\bar{y}_{12} + (-1/4)\bar{y}_{21} + (+1/4)\bar{y}_{22} \quad .$$

It should be noted that the fractional coefficients (within contrasts) are not essential to the valid use of orthogonal comparisons. Rather, it is the ratio of the coefficients to one another (between contrasts) that must be maintained. If the fractions are converted to integers, then the complete collection of linear contrasts may be represented by a contrast matrix as shown in Table 2, where the rows represent the individual cell means while

Insert Table 2 about here

the columns represent the orthogonal contrasts which constitute the elements of the factorial ANOVA design.

Since in the 2 by 2 model: $\alpha_1 = -\alpha_2$, $\beta_1 = -\beta_2$, and $\gamma_{11} = -\gamma_{12} = -\gamma_{21} = \gamma_{22}$, it may be shown that the expected values of each of the Table 2 contrasts contain only the single parameter of interest; that is:

$$E(\hat{\Psi}_A) = E(\bar{y}_{1.} - \bar{y}_{2.}) = E[(\bar{y}_{1.} - \bar{y}_{..}) - (\bar{y}_{2.} - \bar{y}_{..})]$$

$$= E(\hat{\alpha}_1 - \hat{\alpha}_2) = \alpha_1 - \alpha_2$$

$$= 2\alpha_1$$

$$E(\hat{\Psi}_B) = E(\bar{y}_{.1} - \bar{y}_{.2}) = E[(\bar{y}_{.1} - \bar{y}_{..}) - (\bar{y}_{.2} - \bar{y}_{..})]$$

$$= E(\hat{\beta}_1 - \hat{\beta}_2) = \beta_1 - \beta_2$$

$$= 2\beta_1$$

$$E(\hat{\Psi}_{AB}) = E(\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22})$$

$$= E[(\bar{y}_{11} - \bar{y}_{1.} - \bar{y}_{.1} + \bar{y}_{..}) - (\bar{y}_{12} - \bar{y}_{1.} - \bar{y}_{.2} + \bar{y}_{..})$$

$$- (\bar{y}_{21} - \bar{y}_{2.} - \bar{y}_{.1} + \bar{y}_{..}) + (\bar{y}_{22} - \bar{y}_{2.} - \bar{y}_{.2} + \bar{y}_{..})]$$

$$= E(\hat{\gamma}_{11} - \hat{\gamma}_{12} - \hat{\gamma}_{21} + \hat{\gamma}_{22}) = \gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}$$

$$= 4\gamma_{11}$$

so that the contrasts as defined will produce unconfounded estimates of their respective effects.

Thus, if H_{01} is true, then $\Psi_A = 0$. In like manner, $\Psi_B = 0$ and $\Psi_{AB} = 0$ if H_{02} and H_{03} are true. This means that in the 2 by 2 factorial design the hypotheses H_{01} , H_{02}, and H_{03} are equivalent to the hypotheses:

$$H_{01}: \Psi_A = 0, \; H_{02}: \Psi_B = 0, \text{ and } H_{03}: \Psi_{AB} = 0$$

so that hypotheses about equal parameter values are identical to hypotheses about contrasts being equal to zero.

It should be noted that the ANOVA hypotheses H_{01}, H_{02}, and H_{03} written as hypotheses about Ψ_A, Ψ_B, and Ψ_{AB} involve <u>every cell</u> of the 2 by 2 design. As will be seen shortly, the inclusion of every cell (weighted equally) in each contrast is true of all 2^K designs. This is important for a researcher to keep in mind when interpreting significant effects. In such cases, to account for an effect on the basis of anything less than the equal contribution of every cell is to commit a Type IV error.

Interaction in the I by J Design

It should be noted that in the 2 by 2 factorial design there is only one way to define three orthogonal contrasts that relate to the main effect for A, the main effect for B, and their interaction. Thus, the test that the parameter values are equal is identical to the test that the corresponding contrast is equal to zero. As soon as I or J exceeds 2, this last statement is no longer true since there is an infinite number of ways to partition the sum of squares associated with an effect that has three or more levels. This means that the test of equal parameter values does not have a simple counterpart in a test stating that a specified contrast is equal to zero. Instead, the correspondence must be made by a statement relating all possible contrasts as being equal to zero. To illustrate this point, consider a 2 by 3 design, as displayed in Table 3.

Insert Table 3 about here

Since this design consists of six cells, five degrees of freedom are available for partitioning the between groups sum of squares. One possible set of five orthogonal contrasts which may be tested by this factorial design is presented in Table 4[5]. Since I=2, the

Insert Table 4 about here

first column of Table 4 is similar to the first column of Table 2. Differences among the three levels of Factor B are tested by means of two contrasts, each representing one degree of freedom. The contrast in the second column compares Level 1 and Level 2 of Factor B, while the contrast in the third column compares Level 3 of Factor B with Levels 1 and 2 combined. The interaction effects are measured by means of the contrasts defined in the fourth and fifth columns. The contrast in the fourth column is found by multiplying the coefficients of the first column by those of the second, producing a contrast that measures the differential effect of Levels 1 and 2 of Factor B at the two levels of Factor A. Finally, the fifth column is found by multiplying the coefficients of the first and third columns. This contrast measures the differential effect of the combined first two levels and Level 3 of Factor B at the two levels of Factor A.

Any of the five contrasts which were of interest to the researcher could be evaluated as planned comparisons in the manner described in a following section. The important point to note in this discussion is that the interaction contrasts are defined by <u>more than two cells</u> of the design (which will be true for all interaction contrasts), thereby indicating the inappropriateness of attempts to interpret significant interactions strictly on the basis of pairwise or nested statistical comparisons of cell means. Moreover, the expected vslues of both $\hat{\Psi}_{AB_1}$ and $\hat{\Psi}_{AB_2}$ indicate that

they are indeed true interaction contrasts, since:

$$E(\hat{\Psi}_{AB_1}) = E(\bar{y}_{11}) - E(\bar{y}_{12}) - E(\bar{y}_{21}) + E(\bar{y}_{22})$$

$$= (\mu + \alpha_1 + \beta_1 + \gamma_{11}) - (\mu + \alpha_1 + \beta_2 + \gamma_{12})$$

$$- (\mu + \alpha_2 + \beta_1 + \gamma_{21}) + (\mu \pm \gamma_2 + \beta_2 + \gamma_{22})$$

$$= \gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22} \quad \text{and}$$

$$E(\hat{\Psi}_{AB_2}) = E(\bar{y}_{11}) + E(\bar{y}_{12}) - 2E(\bar{y}_{13}) - E(\bar{y}_{21}) - E(\bar{y}_{22})$$

$$+ 2E(\bar{y}_{23})$$

$$= (\mu + \alpha_1 + \beta_1 + \gamma_{11}) + (\mu + \alpha_1 + \beta_2 + \gamma_{12}) - 2(\mu + \alpha_1 \ \beta_3 + \gamma_{13})$$

$$- (\mu + \alpha_2 + \beta_1 + \gamma_{21}) - (\mu + \alpha_2 + \beta_2 + \gamma_{22})$$

$$+ 2(\mu + \alpha_2 + \beta_3 + \gamma_{23})$$

$$= \gamma_{11} + \gamma_{12} - 2\gamma_{13} - \gamma_{21} - \gamma_{22} + 2\gamma_{23}$$

which contain only interaction (γ) parameters.

Since sets of contrasts different from those of Table 4 may be used to partition the individual sums of squares for the B factor and the interaction, it follows that the tests of H_{02}: $\beta_1 = \beta_2 = \beta_3 = 0$ and H_{03}: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{23} = 0$ are not equivalent to the tests of H_{02}: $\Psi_{B_1} = 0$, $\Psi_{B_2} = 0$ and H_{03}: $\Psi_{AxB_1} = 0$,

$\Psi_{AxB_2} = 0$ as was the case in the 2 by 2 design. As soon as I or J exceeds two, the statistical hypothesis of equal parameter values is equivalent to the hypothesis that states that <u>all</u> contrasts are equal to zero. In this sense, H_{02} and H_{03} are equivalent to the hypotheses H_{02}: All $\Psi_B = 0$ and H_{03}: All $\Psi_{AB} = 0$. Because these are the hypotheses that are actually tested in the analysis of variance, one may always use Scheffe'''s method following a rejected overall hypothesis provided that the appropriate type of contrast has been defined, and that the appropriate Scheffe' coefficient is selected. Any number of contrasts may be examined as long as the expected value of the contrast reduces to a contrast in the γ_{ij} only.

Thus, if H_{03}, the overall test of interaction is performed and if the hypothesis is rejected, then Scheffe'''s method will guarantee Type I error protection for all contrasts investigated, but only (<u>a</u>) if they are valid interaction contrasts, and (<u>b</u>) if the Scheffe' coefficient is based on the degrees of freedom associated with the test of H_{03}. Valid interaction contrasts would include tests of interaction effects being different from zero ($\gamma_{ij} = 0$), or from one another ($\gamma_{ij} = \gamma_{i'j'}$), as well as differences between row or column differences ($\Delta_i = \Delta_j$) as described by Marascuilo and Levin (1970). In addition, any contrast in the cell means may be studied, provided that it is shown to be a contrast involving the γ_{ij} only.

This discussion suggests that future textbook writers might well denote analysis of variance F-test hypotheses as H_0: All $\Psi = 0$.

Interactions in 2^K Factorial Designs

Once the general model of a two-factor design is understood, it becomes quite easy to extend the discussion of interactions to more complex designs.

The extension simply involves the addition of more rows and columns to the contrast matrix. For this extension, consider a 2 by 2 by 2 design, as represented in Table 5. In this case, the first, second, and third subscripts refer to the levels of Factors

---

Insert Table 5 about here

A,B, and C respectively. The seven between-group contrast coefficients corresponding to this design are shown in Table 6. Note that, as before, each of the first-order (two-factor) interaction

---

Insert Table 6 about here

---

contrast vectors may be generated by obtaining the products of the constituent factors. Similarly, the A x B x C second-order interaction is the product of the A,B, and C main effect contrast vectors.

In examining the coefficients of the A x B x C interaction more closely, it will be noticed that the first four coefficients represent the B x C interaction for A_1, while the last four coefficients represent the B x C interaction for A_2. The difference between the B x C interaction at Levels A_1 and A_2 may be written as follows:

$$\begin{aligned}\hat{\Psi}_{ABC} &= [(+1)\bar{y}_{111} + (-1)\bar{y}_{112} + (-1)\bar{y}_{121} + (+1)\bar{y}_{122}] \\ &\quad - [(+1)\bar{y}_{211} + (-1)\bar{y}_{212} + (-1)\bar{y}_{221} + (+1)\bar{y}_{222}] \\ &= (+1)\bar{y}_{111} + (-1)\bar{y}_{112} + (-1)\bar{y}_{121} + (+1)\bar{y}_{122} + (-1)\bar{y}_{211} \\ &\quad + (+1)\bar{y}_{212} + (+1)\bar{y}_{221} + (-1)\bar{y}_{222}\end{aligned}$$

which is recognized as the contrast defined in the last column of Table 6. It is easy to show that $\hat{\Psi}_{ABC}$ is an unbiased estimate of

$$\Psi_{ABC} = \gamma_{111} - \gamma_{112} - \gamma_{121} + \gamma_{122} - \gamma_{211} + \gamma_{212} + \gamma_{221} - \gamma_{222} ,$$

a contrast involving the interaction parameters only.

With one-degree-of-freedom tests, the F-test always corresponds to the contrast examined and to its post hoc discussion. Should the three-factor interaction of Table 6 prove to be statistically significant, then the effect is immediately traceable to the equally weighted linear contrast defined in the last column, and not to any isolated cell or cells of the total design. Instead, the

interpretation must include every cell of the design because the test of interaction is identical to the test of H_0: $\Psi_{ABC} = 0$ versus H_1: $\Psi_{ABC} \neq 0$.

Interactions in the I by J by K or Higher-Order Designs

When I,J, or K exceeds 2, then the corresponding analysis of variance hypotheses cannot be stated in terms of a single contrast. In this case, the hypotheses again relate to all possible contrasts that could be generated under the model. Thus, for the three-factor interaction, the classical F-test actually tests the hypothesis H_0: All interaction contrasts are identically equal to zero. The alternative hypothesis is given by H_1: At least one interaction contrast is different from zero.

If H_0 is rejected, confidence intervals may be built around the individual γ_{ijk}, or around linear combinations of the various cells that are indeed true interaction contrasts (not simple comparisons among cell means). Type IV errors are readily avoided by limiting one's verbal interpretation of the interaction to those true interaction comparisons for which the Scheffe′ post hoc confidence interval does not include zero.

Interactions as Planned Comparisons

Some researchers have a number of misconceptions concerning the partitioning of the sum of squares in complex designs. Whereas most know that they may generate one-degree-of-freedom tests "within" a main effect that may be tested as planned orthogonal

comparisons, they fail to realize that a partitioning of the interaction sum of squares is also possible when planned contrasts (in the form of those in Table 4) are specified. The prevailing belief is that while main effects may be decomposed into one-degree-of-freedom contrasts, interactions must be assessed in the context of multi-degree-of-freedom omnibus F-tests. That this is not true is illustrated in the following example.

Consider a cross-sectional study consisting of two factors, Sex (male and female) and Age (6,8,10, and 12 years of age), as portrayed in Table 7. Further suppose that an investigator has

---

Insert Table 7 about here

---

reason to believe that a certain cognitive ability is of such a nature that in the primary grades there is a large sex difference in favor of girls, but that the difference diminishes over the elementary school years. This statement has the flavor of an interaction hypothesis which could be evaluated on a post hoc basis (as outlined in the preceding sections) following the rejection of the hypothesis of no interaction with a statistical test based on three degrees of freedom for the numerator.

But reconsider the investigator's hypothesis. On the surface, it appears that the hypothesis states that the mean profile for the boys over the four age levels is not parallel to the corresponding profile for the girls. Actually, it is more explicit than

that. It states that a relatively large initial girl-boy difference will be observed that will decrease as age increases. If the investigator's hypothesis is correct, then symbolically:

$$(\mu_{11} - \mu_{21}) > (\mu_{12} - \mu_{22}) > (\mu_{13} - \mu_{23}) > (\mu_{14} - \mu_{24})$$

In this case, it is easy to relate the interaction hypothesis to an interaction test for trend using the coefficients for linear, quadratic and cubic components. By referring to a standard table of orthogonal polynomials, as in Hays (1963) or in Kirk (1968), one may use the same coefficients that test for trend within the main effects sum of squares to test the trend interaction hypothesis within the interaction sum of squares. The contrast matrix appropriate for testing this is presented in Table 8.

Insert Table 8 about here

The first column defines the contrast for comparing the girls' and boys' overall (across age) performance. The next three columns constitute three orthogonal contrasts that test for the main effect of age by means of a trend analysis for linear, quadratic, and cubic components. (Depending on the researcher's hypothesis regarding the age main effect, the three trend contrasts would be tested either individually or collectively.) These coefficients are read directly from Table VI of Hays (1963). The last three columns are generated from the first four: Column 5 is found by multiplying the coefficients of Columns 1 and 2 to produce

the linear Sex by Age interaction contrast, Column 6 is the product of Columns 1 and 3, while Column 7 is the product of Columns 1 and 4. These latter two sets of coefficients define the quadratic and the cubic Sex by Age interaction contrasts respectively.

In this example, the contrast of primary interest is defined by the coefficients of Column 5. This contrast is given by:

$$\hat{\Psi}_{SxA(linear)} = (-3)\bar{y}_{11} + (-1)\bar{y}_{12} + (+1)\bar{y}_{13} + (+3)\bar{y}_{14} + (+3)\bar{y}_{21} + (+1)\bar{y}_{22} + (-1)\bar{y}_{23} + (-3)\bar{y}_{24}$$

which may be written as:

$$\hat{\Psi}_{SxA(linear)} = -3(\bar{y}_{11}-\bar{y}_{21}) -1(\bar{y}_{12}-\bar{y}_{22}) +1(\bar{y}_{13}-\bar{y}_{23}) +3(\bar{y}_{14}-\bar{y}_{24})$$

which is seen to have the same basic form as that used for the linear trend for main effects except that the coefficients in this case are applied to the mean sex difference at each of the four age levels. In addition, it should be noted that $\hat{\Psi}_{SxA(linear)}$ is a valid interaction contrast. In the mathematical model, $\hat{\Psi}_{SxA(linear)}$ is an estimate of:

$$\begin{aligned}\Psi_{SxA(linear)} &= -3(\mu + \alpha_1 + \beta_1 + \gamma_{11} - \mu - \alpha_2 - \beta_1 - \gamma_{21}) \\ &\quad -1(\mu + \alpha_1 + \beta_2 + \gamma_{12} - \mu - \alpha_2 - \beta_2 - \gamma_{22}) \\ &\quad +1(\mu + \alpha_1 + \beta_3 + \gamma_{13} - \mu - \alpha_2 - \beta_3 - \gamma_{23}) \\ &\quad +3(\mu + \alpha_1 + \beta_4 + \gamma_{14} - \mu - \alpha_2 - \beta_4 - \gamma_{24}) \\ &= -3(\gamma_{11} - \gamma_{21}) -1(\gamma_{12} - \gamma_{22}) +1(\gamma_{13} - \gamma_{23}) +3(\gamma_{14} - \gamma_{24})\end{aligned}$$

which contains only interaction parameters and, thus, is not confounded with other effects.

Some hypothetical data and an ANOVA based on them may be found in Tables 9 and 10. Calculations of the sums of squares for each source

Insert Table 9 about here

Insert Table 10 about here

of variance and for the planned interaction comparison are as follows:

$$SS_S = Jn \sum_{i=1}^{I} (\bar{y}_{1.} - \bar{y}_{..})^2$$

$$= 4(6)[(18.0 - 15.5)^2 + (13.0 - 15.5)^2] = 24[(-2.5)^2 + (2.5)^2]$$

$$= 24(12.5) = 300$$

$$SS_A = In \sum_{j=1}^{J} (\bar{y}_{.j} - \bar{y}_{..})^2$$

$$= 2(6)[10.0 - 15.5)^2 + (15.0 - 15.5)^2 + (17.5 - 15.5)^2$$

$$+ (19.5 - 15.5)^2]$$

$$= 12[(-5.5)^2 + (-.5)^2 + (2.0)^2 + (4.0)^2]$$

$$= 12(50.5) = 606$$

$$SS_{SxA} = n \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

$$= 6[(14.0 - 18.0 - 10.0 + 15.5)^2 + (6.0 - 13.0 - 10.0 + 15.5)^2$$

$$+ \ldots + (19.0 - 13.0 - 19.5 + 15.5)^2]$$

$$= 6(13) = 78$$

$$\hat{\Psi}_{SxA(linear)} = -3(\bar{y}_{11} - \bar{y}_{21}) -1(\bar{y}_{12} - \bar{y}_{22}) +1(\bar{y}_{13} - \bar{y}_{23}) +3(\bar{y}_{14} - \bar{y}_{24})$$

$$= -3(14-6) -1(18-12) +1(20-15) +3(20-19)$$

$$= -3(8) -1(6) +1(5) +3(1)$$

$$= 22$$

$$SS_{\hat{\Psi}_{SxA(linear)}} = \frac{n(\hat{\Psi}_{SxA(linear)})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}^2} = \frac{6(22)^2}{(-3)^2+(-3)^2+(-1)^2+\ldots+(+3)^2} = \frac{6(484)}{40}$$

$$= 72.6$$

$$SS_{SxA(remainder)} = SS_{\hat{\Psi}_{SxA(quadratic)}} + SS_{\hat{\Psi}_{SxA(cubic)}}$$

$$= SS_{SxA} - SS_{\hat{\Psi}_{SxA(linear)}}$$

$$= 78.0-72.6$$

$$= 5.4$$

Clearly, $\hat{\Psi}_{SxA(linear)}$ gets to the heart of the investigator's query, i.e., whether there exists a decreasing sex difference in the cognitive ability as a function of increasing age. His question is evaluated statistically by weighting the four girl-boy differences by the appropriate coefficients that are related to linear trend. If the investigator were interested in other characteristics of the girl-boy differences, the higher order trend components could be examined individually.

If the factor, Sex: {males, females} were replaced by a factor with more than two levels, e.g.,{Social Class: high, middle, low} , then defining interesting planned orthogonal contrasts might become a little more difficult. However, if the interaction were assessed via the omnibus (with 6 df) test that: All $\psi_{SxSC} = 0$, and if such a test were statistically significant, then contrasts could be defined to compare the mean linear high-low, high-middle, and middle-low differences, using Scheffe´'s procedure. Note that the orthogonality restriction on contrasts is relevant only insofar as partitioning sums of squares into nonoverlapping pieces for hypothesis testing on an a priori basis is concerned. However, if the entire set of contrasts is tested collectively and if such a test produces a significant F, then Scheffe´'s method may be applied to all comparisons -- orthogonal and non-orthogonal alike -- that strike the investigator's fancy, as long as they represent true contrasts among the parameters indicated in the initial test.

It is worth mentioning that the hypothesis of mean girl-boy differences predicts that the differences will decrease as age increases. This is certainly in the mode of a directional hypothesis and therefore to achieve maximum statistical power, it should be analyzed as a directional (one-tailed) alternative. Since the hypothesis is related to a linear contrast, one may perform the test by use of the Student t-distribution, by means of the test statistic:

$$\underline{t} = \frac{\hat{\Psi}_{SxA(linear)}}{S.E._{\hat{\Psi}_{SxA(linear)}}} = \frac{-3(\bar{y}_{11}-\bar{y}_{21})-1(\bar{y}_{12}-\bar{y}_{22})+1(\bar{y}_{13}-\bar{y}_{23})+3(\bar{y}_{14}-\bar{y}_{24})}{\sqrt{\frac{MS_E}{n}\,[(-3)^2+(-3)^2+(-1)^2+(-1)^2+(+1)^2+(+1)^2+(+3)^2+(+3}}$$

which is simply the square root of the $\underline{F}$-ratio based on the same contrast. If the investigator's claim is true, then one would expect that:

$$(\bar{y}_{11}-\bar{y}_{21}) > (\bar{y}_{12}-\bar{y}_{22}) > (\bar{y}_{13}-\bar{y}_{23}) > (\bar{y}_{14}-\bar{y}_{24})$$

which when weighted by the above coefficients would produce a <u>negative</u> value of $\underline{t}$. Thus, as a one-tailed test, the hypothesis H_0: $\Psi_{SxA(linear)} = 0$ should be rejected if the observed $\underline{t} < \underline{t}_{\nu_2}(\alpha)$, where $\underline{t}_{\nu_2}(\alpha)$ is the critical value of $\underline{t}$, based on the degrees of freedom associated with MS_E, at the α (100) percentile.

<u>Interaction in The 2 by 2 Intuitive ANOVA Design</u>

The basic argument presented in this paper is that interaction contrasts defined following a significant $\underline{F}$-ratio must include more than two cells of the design and further, must reduce to a contrast involving the interaction parameters only. Thus, in all of the examples presented, the contrasts examined have been defined in such a way that a linear combination of the cell means was really estimating some linear combination of the γ_{ij}. This was also true for the 2 by 2 design in which it was seen that the only contrast associated with a significant interaction was given by:

$$\hat{\psi}_{AB} = (+1)\bar{y}_{11} + (-1)\bar{y}_{12} + (-1)\bar{y}_{21} + (+1)\bar{y}_{22}$$

$$= (\bar{y}_{11} + \bar{y}_{22}) - (\bar{y}_{12} + \bar{y}_{21})$$

It is worth noting that the form of this contrast is independent of the mathematical ANOVA model and would thereby be encountered under the intuitive ANOVA model discussed by Marascuilo and Levin (1970).

Marascuilo and Levin suggested that many behavioral scientists "intuit" a statistical interaction in much the same way that pharmacists view the joint cumulative effects of two drugs when taken together. For example, neither, either, or both of two drugs (A and B) might be administered to four independent groups of Ss as follows:

Group	Drug Treatment
I	Placebo
II	Drug A
III	Drug B
IV	Drugs A and B

One may represent the intuitive model by means of the factorial design in Table 11. A brief inspection of the four drug treatment combinations

Insert Table 11 about here

indicates why interactions are intuitively traced to a single treatment or cell. In this model, a statistically significant interaction component is immediately attributed to the responses of the subjects in Group IV. However, in order to estimate the magnitude of the interaction, or γ, component at least one comparison would have to be made within the design.

At first glance, it might seem that an estimate of γ could be obtained by comparing the average response of Group IV with the pooled average of Groups II and III by means of the contrast:

$$\hat{\Psi} = \bar{y}_{22} - 1/2(\bar{y}_{12} + \bar{y}_{21})$$

According to the algebra of expected values:

$$E(\hat{\Psi}) = E(y_{22}) - 1/2E(y_{12}) - 1/2E(y_{21})$$

$$= \mu + \alpha + \beta + \gamma - 1/2(\mu + \alpha) - 1/2(\mu + \beta) = + 1/2(\alpha + \beta)$$

Hence, it is clear that this contrast does not do the job, since the interaction effect is partially confounded with the α and β effects.

Alternatively, it might be decided <u>not</u> to average the response of groups II and III, but to evaluate the interaction by means of:

$$\hat{\Psi} = \bar{y}_{22} - (\bar{y}_{12} + \bar{y}_{21})$$

which in this case provides an unbiased estimate of:

$$E(\hat{\Psi}) = \gamma - \mu$$

Unfortunately, this estimate of γ is biased, in that it tends to underestimate the effect of γ by the amount μ. Moreover, the linear combination considered is not a legitimate contrast since the sum of the coefficients does not add to zero. However, it can be modified to form a contrast by considering the sum of the averages in Groups I and IV as contrasted with the sum of the averages in Groups II and III. With the contrast:

$$\hat{\Psi} = (\bar{y}_{11} + \bar{y}_{22}) - (\bar{y}_{12} + \bar{y}_{21})$$

it is seen that:

$$E(\hat{\Psi}) = (\mu + \mu + \alpha + \beta + \gamma) - (\mu + \alpha + \mu + \beta) = \gamma$$

Clearly, this is the contrast that is appropriate for determining the magnitude of the interaction effect. It involves all four cells of the design in exactly the same manner as suggested when the 2 by 2 ANOVA design was discussed in the earlier sections of this paper. Thus, it is readily apparent that even though one may subscribe in principle to the intuitive interaction model, in practice when it comes to estimating and isolating the interaction effects. even contrasts among the parameters of the intuitive model reduce to exactly the same contrasts encountered in the mathematical ANOVA model.

The meaning of this entire discussion on Type IV errors manifested by interactions in ANOVA designs should be clear for the behavioral scientist. Significant interactions examined as either planned or post hoc comparisons must be evaluated either in terms of the interaction parameters of the model or in terms of cell means that define contrasts that reduce to comparisons among the interaction parameters of the model. If it is seen that the expected value of a contrast defined in terms of cell means contains any α, β, or μ of the design, then it is immediately known that the contrast is

not a valid interaction contrast and should therefore not be discussed as though it were related to a significant interaction component. Just as tests of interactions are orthogonal to tests of main effects, interaction contrasts are orthogonal to main effect contrasts and therefore their expected values are independent of one another. If these principles are kept in mind and if each interesting interaction contrast is inspected in terms of its expected value, then Type IV errors in the interaction model should, like old soldiers, fade away.

References

Davis, D.J. Flexibility and power in comparisons among means. Psychological Bulletin, 1969, 71, 441-444.

Dixon, W.J., and Massey, F.J., Jr. Introduction to Statistical Analysis. McGraw-Hill: New York, 1969.

Dunn, O.J. Multiple comparisons among means. Journal of The American Statistical Association, 1961, 5, 52-64.

Dunn, O.J. Multiple comparisons using rank sums. Technometrics, 1964, 6, 241-252.

Goodman, L.A. Interactions in multidimensional contingency tables. Annals of Mathematical Statistics, 1964, 35, 632-646.

Hays, W.L. Statistics for Psychologists. Holt, Rinehart, and Winston: New York, 1963.

Hummel, T.J., and Sligo, J.R. Empirical comparisons of univariate and multivariate analysis of variance procedures. Psychological Bulletin, in press.

Kirk, R.E. Experimental Design for the Behavioral Sciences. Brooks-Cole: Belmont, California, 1968.

Marascuilo, L.A. Large-sample multiple comparisons. Psychological Bulletin, 1966, 65, 280-290.

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

Marascuilo, L.A. and McSweeney, M. Nonparametric post hoc comparisons for trend. Psychological Bulletin, 1967, 67, 401-412.

Mendenhall, W. The Design and Analysis of Experiments. Brooks-Cole: Belmont, California, 1968.

Miller, R.G., Jr. Simultaneous Statistical Inference. McGraw-Hill: New York, 1966.

Morrison, D.F. Multivariate Statistical Methods. McGraw-Hill: New York, 1967.

Petrinovich, L.F. and Hardyck, C.D. Error rates for multiple comparison methods: Some evidence concerning the frequency of erroneous conclusions. Psychological Bulletin, 1969, 71, 43-54.

Scheffe´, H. A method for mudging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Scheffe´, H. The Analysis of Variance. John Wiley and Sons: New York, 1959.

Steel, R.G.D. Some rank sum multiple comparison tests. Biometrika, 1961, 71, 539-552.

Timm, N.H. Multivariate Analysis in Educational and Psychological Research. In preparation.

Tukey, J.W. The problem of multiple comparisons. Ditto, Princeton University, 1953.

Table 1. Two-Factor Design with I=2 and J=2 in Terms of the Observed Means.

Factor A B	A_1	A_2	Mean
B_1	$\bar{y}_{11}$	$\bar{y}_{21}$	$\bar{y}_{.1}$
B_2	$\bar{y}_{12}$	$\bar{y}_{22}$	$\bar{y}_{.2}$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{..}$

Table 2. Contrast Matrix for Partitioning the Sum of Squares in a 2 by 2 Factorial Design into Three Orthogonal Components Related to Main Effect for A, Main Effect for B, and their Interaction.

Cell Mean	Contrast		
	$\hat{\psi}_A$	$\hat{\psi}_B$	$\hat{\psi}_{AB}$
$\bar{y}_{11}$	1	1	1
$\bar{y}_{12}$	1	-1	-1
$\bar{y}_{21}$	-1	1	-1
$\bar{y}_{22}$	-1	-1	1

Table 3. Two-Factor Design with J=2 and J=3 in Terms of the Observed Means

Factor A B	A_1	A_2	Mean
B_1	$\bar{y}_{11}$	$\bar{y}_{21}$	$\bar{y}_{.1}$
B_2	$\bar{y}_{12}$	$\bar{y}_{22}$	$\bar{y}_{.2}$
B_3	$\bar{y}_{13}$	$\bar{y}_{23}$	$\bar{y}_{.3}$
Mean	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{..}$

Table 4. Contrast Matrix for Partitioning the Sum of Squares in a 2 by 3 Factorial Design into Five Orthogonal Components related to Main Effect for A, Main Effect for B, and their Interactions.

Cell Mean	Contrast $\hat{\Psi}_A$	$\hat{\Psi}_{B_1}$	$\hat{\Psi}_{B_2}$	$\hat{\Psi}_{AxB_1}$	$\hat{\Psi}_{AxB_2}$
$\bar{y}_{11}$	1	1	1	1	1
$\bar{y}_{12}$	1	-1	1	-1	1
$\bar{y}_{13}$	1	0	-2	0	-2
$\bar{y}_{21}$	-1	1	1	-1	-1
$\bar{y}_{22}$	-1	-1	1	1	-1
$\bar{y}_{23}$	-1	0	-2	0	2

Table 5. Three-Factor Design with I=2, J=2, and K=2 in Terms of the Observed Means

Factor A / B / C	A_1		A_2	
	B_1	B_2	B_1	B_2
C_1	$\bar{y}_{111}$	$\bar{y}_{121}$	$\bar{y}_{211}$	$\bar{y}_{221}$
C_2	$\bar{y}_{112}$	$\bar{y}_{122}$	$\bar{y}_{212}$	$\bar{y}_{222}$

Table 6. Contrast Matrix for Partitioning the Sum of Squares in a 2 by 2 Factorial Design into Seven Orthogonal Components Related to Main Effects for A, B, and C and their Interactions.

Cell Mean	Contrast $\hat{\psi}_A$	$\hat{\psi}_B$	$\hat{\psi}_C$	$\hat{\psi}_{AB}$	$\hat{\psi}_{AC}$	$\hat{\psi}_{BC}$	$\hat{\psi}_{ABC}$
$\bar{y}_{111}$	1	1	1	1	1	1	1
$\bar{y}_{112}$	1	1	-1	1	-1	-1	-1
$\bar{y}_{121}$	1	-1	1	-1	1	-1	-1
$\bar{y}_{122}$	1	-1	-1	-1	-1	1	1
$\bar{y}_{211}$	-1	1	1	-1	-1	1	-1
$\bar{y}_{212}$	-1	1	-1	-1	1	-1	1
$\bar{y}_{221}$	-1	-1	1	1	-1	-1	1
$\bar{y}_{222}$	-1	-1	-1	1	1	1	-1

Table 7. Two-Factor Design of Sex by Age.

Factor Age \ Sex	Girls	Boys
6 years	$\bar{y}_{11}$	$\bar{y}_{21}$
8 years	$\bar{y}_{12}$	$\bar{y}_{22}$
10 years	$\bar{y}_{13}$	$\bar{y}_{23}$
12 years	$\bar{y}_{14}$	$\bar{y}_{24}$

Table 8. Contrast Matrix for Table 7 Based on Tests for Linear, Quadratic, and Cubic Trends.

Cell Mean	Contrast						
	$\hat{\Psi}_S$	$\hat{\Psi}_{A(linear)}$	$\hat{\Psi}_{A(quad.)}$	$\hat{\Psi}_{A(cubic)}$	$\hat{\Psi}_{SxA(linear)}$	$\hat{\Psi}_{SxA(quad.)}$	$\hat{\Psi}_{SxA(cubic)}$
$\bar{y}_{11}$	1	-3	1	-1	-3	1	-1
$\bar{y}_{12}$	1	-1	-1	3	-1	-1	3
$\bar{y}_{13}$	1	1	-1	-3	1	-1	-3
$\bar{y}_{14}$	1	3	1	1	3	1	1
$\bar{y}_{21}$	-1	-3	1	-1	3	-1	1
$\bar{y}_{22}$	-1	-1	-1	3	1	1	-3
$\bar{y}_{23}$	-1	1	-1	-3	-1	1	3
$\bar{y}_{24}$	-1	3	1	1	-3	-1	-1

Table 9. Hypothetical Performance on a Cognitive Task, by Boys and Girls at Four Age Levels

Factor Sex / Age	Girls	Boys	Across Sex
6	14	6	10.0
8	18	12	15.0
10	20	15	17.5
12	20	19	19.5
Across Age	18.0	13.0	15.5

Note: There are 6 Ss per cell (n=6), and the mean square error (MS_E) associated with these data is 16.0.

Table 10. Analysis of Variance Table for the Data in Table 9, including a Planned Interaction Contrast.

Source	df	SS	MS
Sex	1	300	300
Age	3	606	202
Sex by Age	3	78	
$\hat{\Psi}_{SxA(linear)}$	1	72.6	72.6
$\hat{\Psi}_{SxA(remainder)}$	2	5.4	2.7
Error	40	640	16.0

Note: The sums of squares are based on the means in Table 9, which are proportional to those obtained with Table 8's coefficients.

Table 11. The Intuitive 2 by 2 Design

Factor Drug A Drug B	No	Yes
No	(I) μ	(II) $\mu + \alpha$
Yes	(III) $\mu + \beta$	(IV) $\mu + \alpha + \beta + \gamma$

Footnotes

[1]Readers who are still wondering about Type III errors may refer to assorted definitions reviewed by Marascuilo and Levin (1970).

[2]The authors are grateful to Professors Maryellen McSweeney, Neil H. Timm, and M.I. Charles E. Woodson for reading an initial draft of this paper, and recommending several helpful modifications.

[3]"Appropriate" is used advisedly here, in the sense that the Scheffe' procedure is the only procedure that corresponds exactly to the initial test of hypothesis. Whether Scheffe''s procedure is desirable (with respect to statistical power, for example) is another issue which has been discussed elsewhere (e.g., Petrinovich and Hardyck, 1969).

[4]Although the discussion and examples throughout this paper will be based on the assumption of equal cell n's, the same general principles may be extended to designs with unequal cell frequencies.

[5]Note that even though the two contrasts for the B factor are orthogonal in this case, the orthogonality restrictions of a factorial design apply only to between source (i.e., main effects and interaction) sets of contrasts. For a more comprehensive treatment of contrasts and the general linear model, Mendenhall's (1968) book is an excellent source.