Misinterpretation of Interaction Effects: A Reply to Rosnow and Rosenthal

Donald L. Meyer
University of Pittsburgh

The structural model of analysis of variance for multiway tables, with main effects and interaction parameters, is overspecified. Only the set of estimable functions of the cell means is useful. The individual estimates of the parameters are artificial as regards the underlying scientific process. Rosnow and Rosenthal (1989a) stated that it is "absolutely necessary" (pg 146) that interactions should be interpreted by examining the usual estimates of the interaction parameters. This is incorrect.

Rosnow and Rosenthal (1989a) stated that it is an error to report cell means to interpret the pattern of an observed interaction. Because I suspect that several people will be concerned with this statement, especially when Rosnow and Rosenthal (1989b) stated that interaction effects are "the universally most misinterpreted empirical results in psychology" (p. 1282), I felt this reply was in order. As an aside, I find it fascinating that a four-page article was published explaining how to interpret four numbers—more than 50 years after R. A. Fisher (1935) explained the analysis of variance (ANOVA).

Reviewers of a previous version of this article called my attention to a number of articles also discussing interaction effects in the ANOVA: Marascuilo and Levin (1970, 1976), Levin and Marascuilo (1972, 1973), Games (1973), and Boik (1979). The same lack of sensitivity to the principal point of the present article is evident in some of these articles.

Rosnow and Rosenthal (1989a) present an "algorithm" for obtaining two-way interaction effects "corrected" for main effects. They are seemingly unaware that their algorithm is nonestimable. The model is overspecified. Only the set of estimable functions of the cell means is useful. The usual constraints specify that the row and column estimates both sum to zero and the interaction estimates give a total of nine parameters (effects). It is at once obvious that four observed cell means cannot estimate nine effects!

How then is progress made? The answer is that certain functions of the parameters can be estimated. In particular, the difference of the differences of the cell means in row 1 and the means in row 2 estimates the corresponding difference of the differences of the $\tau(ij)$:

$$[m(1, 2) - m(1, 1)] - [m(2, 2) - m(2, 1)].$$

This is the familiar interaction contrast shown in statistics textbooks.

For concreteness, consider the example used by Rosnow and Rosenthal (1989a) of an investigator studying the despair of bereavement of family members when a child dies. Grief intensity is the dependent variable; health and sex of child are the two "factors."

At the top of Table 1 is a hypothetical set of true population means consistent with Rosnow and Rosenthal's (1989a) ranking of the cells: Healthy male > healthy female > unhealthy female > unhealthy male. Next is presented the "scientific" effects, or what Levin and Marascuilo (1973) refer to as the "latent structure of the variable" (p. 308). This structure specifies that the true means arise from combinations of additive row, column, and interaction effects, as in Model 1. These "true" effects show that being healthy adds 3 points to bereavement, and being unhealthy adds 10 points. Girls receive 5 points; boys receive 0. The numbers in the cells are the interaction effects when sex is mixed with health status. Healthy boys have a true mean of 17 because they are in a cell receiving 3 for healthy, 0 for boys, 12 for the interaction, and 2 for the grand mean. Similarly, the other three true means result as shown. These effects were simply made up and are just one of the infinite number of sets of possible effects consistent with the true means.

Samples from the four groups are taken, and suppose our observed cell means are the same as the true means: a successful research study! The bottom of Table 1 shows the usual estimates using Model 1. These are calculated using constraints on the estimates of the parameters. I emphasize this to show that the statistical analysis is arbitrary as regards the actual values of the parameters. The usual constraints specify that the row and column estimates both sum to zero and the interaction estimates...
calculate the variance of the interaction contrast and, if desired, the variance of these simple contrasts.

The example also shows that even the difference and order of row and column effects are misestimated! The true scientific effects show a 7-point effect favoring unhealthy, but the ANOVA model shows a difference of 10 points favoring healthy. When Rosnow and Rosenthal “peel” away main effects, they use unreal main effects. It is a well-known fact that when interaction occurs, main effects are not readily interpretable. Rosnow and Rosenthal give a second example in which they calculate meaningless F-tests for main effects in the presence of a significant interaction. At least they are in good company, because this is seen repeatedly both in articles and textbooks!

This second example shows how Rosnow and Rosenthal’s advice can confuse even themselves. The example concerned inexperienced (I) and experienced (E) ball players assigned to either a control treatment or a “Ralphing” treatment. Both groups scored 3 in the control condition, and I went up to 5 and E went up to 7 in the Ralphing condition. Both groups benefited from Ralphing, but the experienced group benefited even more. Rosnow and Rosenthal plotted the interaction estimates and obtained the +, - crossing pattern. They then said, “The experienced ball players benefited moderately from Ralphing to the same degree that inexperienced ball players were harmed by it” (Rosnow & Rosenthal, 1989a, p.146).

The theoretical cell means are estimable by the observed cell means (or more generally, by linear functions of the means). In plain words, they reflect reality. When interaction occurs, they are the best guesses (non-Bayesian) of future behavior for the various factor combinations defining the cells. Furthermore, in the presence of interaction, the structural model with main effects is of no use. Rosnow and Rosenthal (1989a) state, “The point of this article is to emphasize that if investigators are claiming to speak of an interaction, the exercise of looking at the corrected cell (or condition) means is absolutely necessary” (p.146). This is misguided. The usual graph of cell means shows how the two factors are behaving together. The interaction is seen in the lack of parallelism of the two lines. Rosnow and Rosenthal would have us always plot the lines as crossing!

Marascuilo and Levin (1970) stated that many of the errors in analysis after rejection of a test for interaction come from “an incorrect understanding that researchers have concerning interaction in a factorial design” (p. 414). They called this the “intuitive model,” or “synergistic model.” Their example concerns four groups: placebo, Drug A, Drug B, both Drug A and B. Their model specifies the following:

<table>
<thead>
<tr>
<th>Group</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>placebo</td>
<td>M</td>
</tr>
<tr>
<td>Drug A</td>
<td>M + A</td>
</tr>
<tr>
<td>Drug B</td>
<td>M + B</td>
</tr>
<tr>
<td>Drug A, B</td>
<td>M + A + B + (AB)</td>
</tr>
</tbody>
</table>

Marascuilo and Levin (1970) identify the (AB) parameter as the interaction resulting from the synergistic effect of the two drugs administered together. They go on to say that in this model it makes good sense to think of interaction in a single cell, but in...
the generally applied ANOVA model, an interaction is “a component that involves every cell of the design and not just the cell representing the joint administration of the two drugs” (2 × 2 designs; Marascuilo & Levin, 1970, p. 416).

Table 2 shows the estimates using the same true means as in the top of Table 1. Lo and behold! The synergistic interaction parameter is estimated to be 4, which is the value of the interaction contrast of the cell means calculated previously. Again the main effects of drugs, A = -1 and B = 8, bear no resemblance to either the usual estimates or the “latent” effects of Table 1. This “intuitive” model is just another way of parameterizing four cell means. Whether the true science is synergistic or follows the latent structural model is unknowable from four data points (or more in a larger two-way table!)

Much of the discussion is valuable because Marascuilo and Levin (1970) emphasize interaction as the difference of differences and they address the problem of multiple-error rates. However, they seemed to have not followed their own dictum that interaction involves every cell when they made a rather large issue of calculating confidence intervals for single interaction parameters. Games made almost the same criticism when he pointed out that the interaction parameters “have no generalizability to a modified replication” (Games, 1973, p. 305). As I have shown, they do not even have utility for a current data set, except as they enter estimable functions.

Most presentations of the structural model (with more parameters than data points) in statistics textbooks for the behavioral sciences are incomplete. This is probably because of the desire to avoid rather heavy mathematics. However, it is unfortunate that the distinction between the original (scientific) model and the reparameterized model is blurred. The ANOVA is an analysis of estimable functions, not of individual parameters. Two important points are (a) any reparameterization leads to the same estimate of an estimable function, and (b) it is possible to test hypotheses only about estimable functions (Kempthorne, 1952). The failure to make the distinction has misled many, the most recent being Rosnow and Rosenthal (1989a). Once one thinks in terms of cell means or the cell means model as Marascuilo and Levin did later (Marascuilo & Levin, 1976), most of the issues disappear, because contrasts of the cell means are estimable functions. In a 2 × 2 table, four means can be compared in 11 different ways using unit weights. Only one of those ways directly addresses interaction: the difference of the differences. If a researcher wants to report other comparisons like “simple effects” and if those comparisons are deemed useful in interpretation, then by all means they should be reported. I believe Games (1973) subscribed to that view. If one wants to analyze various interaction contrasts in larger tables, refer to textbooks such as Kirk (1982) and the article by Boik (1979). The only real issue is not how to interpret interactions, but how one controls significance levels in multiple-testing situations.

The structural model with only main effects is still overparameterized but seemingly causes no difficulties in interpretation.

To all those who make graphs of cell means to show the interaction as nonparallel lines, ignore the Rosnow and Rosenthal articles and keep on plotting!

References


Table 2

<table>
<thead>
<tr>
<th>Drug B</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>M + B</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>M + A</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

| Yes | M + A + B + (AB) = 6 - 1 + 8 + 4 |
| No  | M = 6                           |

Received November 29, 1989
Revision received October 4, 1990
Accepted December 24, 1990