ON THE RELATIVE POWER OF INTERACTION ANALYSIS

Richard L. Tate
Florida State University

ABSTRACT

It is argued that statements in the current literature suggesting that interaction effects are, in general, as easy to detect as main effects are misleading. Different effect definitions which produce different conclusions about the relative power of interaction analysis are considered for both factorial ANOVA and aptitude-treatment-interaction models. Based on what is defined as a reasonable specification of "comparable" effects, it is concluded that the power for simple main and interaction effects is, in general, lower than that for the analysis of main effects.

INTRODUCTION

Several statements in the current literature about the relative power of interaction analysis are potentially misleading. Some seem to argue that the power for the detection of interactions in a study is, in general, the same as the power for main effects. For example, Cronbach and Snow (1977, p. 81) say, in reaction to statements in the first edition of Cohen (1969, 1977), "There is no 'relative weakness of interaction tests' and the weakness does not 'progress sharply with higher orders'." This conclusion is supported by reference to a 2x2x2 factorial ANOVA design, where the assumption of equal interaction and main effects (using the classical or deviation definition of effects) leads to the conclusion that there is equal power for the detection of the interaction and main effects. Cohen (1977), in the second edition of his book, also notes

(p. 374) that the power for equal interaction and main effects (again using the deviation definition of effects) in 2^k factorial ANOVA is equal. Cronbach and Snow also argue (p. 81) that the same conclusion holds true for the regression approach to aptitute-treatment-interaction (ATI) analysis because "the power of detecting a β of a certain size is the same no matter whether it is for a main effect or an nth-order interaction."

On the other hand, Cronbach and Snow (p. 46) also emphasize that much larger sample sizes are needed to detect the interactions in ATI analysis compared to the sample sizes typically used for main effects research. They explain this by noting (p. 82) that the interaction which they feel to be of practical importance is a much smaller effect than the corresponding "threshold" main effect size. They assume that a correlation between the dependent variable and the aptitude of 0.4 (corresponding to the aptitude main effect) is of practical importance; they then reason further that it would be desirable to detect when there is a difference in the correlations for the two treatments corresponding to a zero correlation for one treatment and a 0.4 correlation for the other.

The juxtaposition by Cronbach and Snow of apparently contradictory discussions about the relative power for interactions seems to suggest that their conclusion about ATI's (i.e., that they are difficult to detect) is viewed as an exception to a general rule that there is no relative power weakness for interaction analysis. The purpose of this paper is to suggest that Cronbach and Snow's conclusion about ATI's also applies, in general, to all interaction analysis, including that associated with factorial ANOVA. First, there is a reminder that a definition of "comparable" effects is necessary in order to draw any conclusions about relative power. Then, power comparisons for main, simple main, and interaction effects are discussed for factorial ANOVA and ATI models. It is illustrated that conclusions about the power for interactions compared to that for main effects, assuming equal effects, depends on whether deviation, difference, or explained variance definitions of the effects are used. Finally, it is argued that, under reasonable specification of comparable effects, there is in general relatively lower power for interaction analysis, and that this weakness does "progress sharply with higher orders."

SOME POWER COMPARISONS

Since the factorial ANOVA model and the ATI model are both special cases of a general regression model, the power formulation and tables presented by Cohen (1977) for regression are used here. The required parameters are the significance level, the hypothesis degrees of freedom, and the noncentrality parameter, $L = f^2 df_e$, where the effect size parameter f^2 is $\Delta R^2/(1-R^2)$, R^2 is the coefficient of determination for the full model, ΔR^2 is the increment in R^2 due to the hypothesis, and df_e is the error degrees of freedom. When a hypothesis can be represented by a single independent variable, the noncentrality parameter can also be expressed by $L = b_j^2/S_b^2$, where b_j is the partial regression coefficient for the independent variable and S_b is the associated standard error. This alternative expression is based on the equivalence of the t-test of a b_j in a linear model and the F-test of the ΔR^2 for the independent variable, given all other variables.

Factorial ANOVA. There are three relatively common ways of defining the effects in a factorial ANOVA. Consider, for example, a simple 2x2 design with factors A and B. In the most common classical or deviation definition of effects, each main effect is just the deviation of a marginal mean from the grand mean; for example, the main effect for the i^{th} level of factor A, α_i, is defined as $(\mu_i - \mu..)$. A simple main effect is the deviation of a cell mean from either the associated A or B marginal mean; for example, the simple main effect of the i^{th} level of A in the j^{th} level of B, $(\alpha_i)_j$, is defined as $(\mu_{ij} - \mu._j)$. Finally, the interaction effect for the ij^{th} cell, γ_{ij}, is the deviation of the cell mean from what would be predicted using only the additive model, i.e., $(\mu_{ij} - (\mu.. + \alpha_i = \beta_j))$ where β_j is the main effect for the j^{th} level of B. For the 2x2 design, the two deviation main effects for each factor are equal in magnitude and the four interaction effects are equal in magnitude. The deviation effects for this simple design can be obtained using the regression approach with 1, -1 coding.

A second definition, called here the difference definition, is based on pairwise comparisons. Thus, for the 2x2 design, the only main effect for factor A is $(\delta_a = \mu_{1.} - \mu_{2.})$, the only simple main effect for A in the j^{th} level of B is $((\delta_a)_j = \mu_{1j} - \mu_{2j}))$, and the only interaction effect is the difference of simple main effects, i.e.,$(\delta_{ab} = (\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22}))$. The difference effects can be obtained with the regression approach using 0,1 coding. A third way to define ANOVA effects is to use the explained variance which is associated with each effect. Thus, for the current comparison, the effect will be defined as $f^2 = \Delta R^2/(1-R^2)$ for the main and interaction effects. (This definition will not be used for the simple main effects here.)

The relationships among these three effect definitons are simple. In the 2x2 design, the difference main effect is twice as large as the deviation main effects for each factor, the difference interaction is four times as large as the deviation interactions, and the difference simple main effects are twice as large as the corresponding deviation simple main effects. The relationship between the deviation and difference definitions and the f^2 explained variance definition is represented in the expression $L = f^2 df_e = b_j{}^2/S_b{}^2$.

Consider now the power for the interaction effect and the simple main effect relative to that for the main effect under each definition. For the deviation and difference definitions, the $L = b_j^2/S_b^2$ definition will be used. Thus, assuming appropriate coding in the regression approach, the ratio of the L for the simple main or interaction effect to that for the main effect will be

$$L/L_{me} = (b/b_{me})^2(S_{me}/S)^2.$$

(This expression indicates that the Cronbach and Snow, 1977, p. 81, statement cited above, that there is equal power for equal β's in an ATI model, is true only when the corresponding standard errors are also equal.) Assuming equal cell size of n, the standard error of estimate can be obtained for any effect in a factorial ANOVA by using

$$S = \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\Sigma\ c_k^2}$$

where $\hat{\sigma}$ is the standard error of estimate, and the effect of interest is equal to $\Sigma c_k \mu_k$ (where the sum is over all cells in the design). The L ratio for the variance definition is simply $L/L_{me} = f^2/f_{me}^2$. The resulting L ratios for the simple main and interaction effects under the deviation, difference, and variance definitions are shown for the 2x2 design in the upper portion of Table 1.

Any conclusion about the relative power for different effects will require a definition of "comparable" effects, i.e., a definition of the threshold effect size for each effect which is considered to be of practical importance. One apparently reasonable definition would be to assume equal effects. The resulting power under equal effects is also shown in Table 1 for the three effect definitions. It has been assumed for Table 1 that the significance level is .05 and the sample size is that required to provide a power of .90 for the main effect.

The deviation and difference definitions produce the same conclusion for the simple main effect; under the assumption of

Table 1

Relative Power for Interaction Analysis with Factorial ANOVA and ATI Models

Effect	L Ratio[a]			Power[b]		
	Dev	Diff	Var	Dev	Diff	Var
2x2 ANOVA and ATI						
Main	1	1	1	.90	.90	.90
Simple main	(1/2)ERS	(1/2)ERS	-	.63	.63	-
Interaction	(1)ERS	(1/4)ERS	(1)ERS	.90	.37	.90
2x2x2 ANOVA						
Main	1	1	1	.90	.90	.90
Simple main	(1/2)ERS	(1/2)ERS	-	.63	.63	-
Simple simple main	(1/4)ERS	(1/4)ERS	-	.37	.37	-
(Global) two-way interaction	(1)ERS	(1/4)ERS	(1)ERS	.90	.37	.90
Simple two-way interaction	(1/8)ERS	(1/8)ERS	-	.21	.21	-
Three-way interaction	(1)ERS	(1/16)ERS	(1)ERS	.90	.13	.90

[a] This is the ratio of the noncentrality L parameter for the effect of interest to that for the main effect. The effect definitions are the deviation ("Dev"), difference ("Diff"), and explained variance ("Var") definitions. "ERS" is the effects ratio squared for each effects definition.

[b] The power for each effect is computed assuming that the ERS is equal to one, the significance level is .05, and the sample size is that necessary to provide a power of .90 for the main effect (i.e., $L_{me} = 10.15$).

equal effects, the L ratio is equal to 1/2, resulting in a power of .63 for the simple main effect compared to the power of .90 for the main effect. The three effect definitions result in different conclusions, however, about the relative power for interactions. For the deviation and variance definitions, the L ratio is equal to one and it is concluded that the power for interactions is equal to that for main effects. In contrast, the L ratio is equal to 1/4 under the difference definition, resulting in a much lower power of .37 for the interaction.

The reason that different effect definitions produce different conclusions about the relative power for interactions follows from the relationships among the effects under different definitions which were noted above. It is easy to show

$$(\alpha_i)_j/\alpha_i = (\delta_a)_j/\delta_a$$

and

$$\gamma_{ij}/\alpha_i = (1/2)(\delta_{ab}/\delta_a) = f^2_{ab}/f^2_a$$

The ratio of simple main effect to main effects is equal for the deviation and difference definitions; therefore, both definitions produce the same conclusion about the relative power of simple main effects. On the other hand, the inequality among the interaction to main effect ratio indicates that the assumption of equal effects under the difference definition corresponds to a relatively smaller interaction than the interaction which is implied by assuming equal effects under the deviation and variance definitions. Thus, the power for the interaction under the difference definition is lower. This can be shown more explicitly by considering the ratio of the L ratios for the deviation and difference definitions. When it is recognized that the relationship between the standard errors for an effect under the two definitions corresponds to the relationship between effects noted earlier, it can be shown that

(L/L_{me})diff$/(L/L_{me})$dev is equal to 1/4.

The inequalities among the interaction to main effect ratios also have implications for judgments of practical significance. For example, the practical importance of interactions is commonly judged on the basis of ΔR^2, apparently using the same criterion or threshold value as used for main effects. If, however, it is decided that an appropriate definition of comparable effects corresponds to equal effects under the difference definition, then an interaction ΔR^2 only one half of the threshold value for the main effect would be of practical significance.

Similar conclusions will hold for higher order factorial designs. The L ratios and power (under the assumption of equal effects) for the simple main, simple simple main, (global) two-way interaction, simple two-way interaction, and three-way interaction effects for a 2x2x2 design are shown for the three effect definitions in the lower portion of Table 1. As before, the power for the simple main and simple simple main effects is lower than that for the main effect under both the deviation and difference definitions. Also, the power for the global two-way interaction and three-way interaction is equal to that for the main effect under the deviation and variance definitions. This corresponds to the statement of Cronbach and Snow (1977, p. 81) cited above, "There is no 'relative weakness of interaction tests' and weakness does not 'progress sharply with higher orders'." However, as before there is lower power for the interactions under the assumption of equal effects in the difference definition. In addition, the power for the simple two-way interaction is lower under both the deviation and difference definitions.

A striking feature of the comparison of the 2x2 and the 2x2x2 results in Table 1 is the decrease in power associated with the higher order design. The power for the simple simple main effect in the 2x2x2 design is only .37 under the deviation and difference definitions. Also, if equal effects under the difference definition are assumed, the power for the three-way interaction is only .13.

Aptitude-Treatment-Interaction. ATI effect definitions which are analogous to the deviation, difference, and explained variance definitions for factorial ANOVA produce similar comparisons and conclusions. Suppose an ATI model has a single categorical treatment variable with two levels represented by the coded variable X_1, a single aptitude variable, X_2, and the interaction term $X_3 = X_1X_2$. Deviation effects are obtained with 1,-1, coding, for X_1, while 1,0 coding produces difference effects. The resulting models for the two treatment levels are:

Deviation effects:

$$\mu_1 = (b_0 + b_1) + (b_2 + b_3)X_2$$

$$\mu_2 = (b_0 - b_1) + (b_2 - b_3)X_2$$

Difference effects:

$$\mu_1 = (b_0' + b_1') + b_2' + b_3')X_2$$

$$\mu_2 = b_0' + b_2'X_2$$

The interpretation of the model parameters follows from these expressions. For example, the interaction parameter under the deviation definition, b_3, is the deviation of the aptitude slope for a treatment from a mean slope, while the same parameter under the difference definition, b_3', is the difference in the slopes for the two treatments. Thus, b_3' is twice as large as b_3.

Consider the power for the simple main effect of the aptitude and the aptitude-treatment-interaction relative to that for the main effect of the aptitude under an additive model. The standard errors for each of these effects derived under the difference definition, assuming equal means and variances of aptitude in the two treatments, are

$$S_{me} = (\hat{\sigma}/\sigma_a) \sqrt{1/(2n - 1)}$$
$$S_{sme} = (\hat{\sigma}/\sigma_a) \sqrt{1/n}$$
$$S_{ati} = (\hat{\sigma}/\sigma_a) \sqrt{(2n - 1)/(n^2 - n)}$$

where σ_a is the within-treatment aptitude standard deviation and n is the sample size for each treatment. When the square of the ratio of standard errors is formed for both the simple main and interaction effects, the resulting L ratios closely approximate those shown for the difference definition in Table 1 for the 2x2 ANOVA design. Thus, under the assumption of equal difference effects, the relative power for these effects is also that shown in Table 1. That is, the power for the simple main effect of the aptitude is .63 and that for the ATI is .37 compared to the power of 0.90 for the main effect of the aptitude under the additive model. In the same fashion, the deviation and variance definitions result in L ratios and powers which are approximately (for the deviation effects) or exactly (for the variance effects) equal to those for the corresponding 2x2 ANOVA results. Thus, the same conclusions apply for the 2x2 ANOVA effects and the ATI effects considered.

CONCLUSIONS

Clearly, it is necessary to use some care in the definition of comparable or threshold effects which is required for statements about relative power. When equal effects are assumed, the power comparisons above illustrate that the conclusion about the relative power for interaction effects will depend on the apparently arbitrary choice of an effects definition. If the deviation effects are used, it will be concluded that the power for the detection of interaction effects is equal to that for main effects. The same conclusion would be reached using the variance definition. On the other hand, if the researcher prefers and uses difference effects, it will be concluded that there is a relative weakness for interaction effects. The point here is not that it is really important which effects definition is used, but that a researcher should not casually assume, without reflection, that "comparable" effects imply equal effects under the effects definition being used.

Since conclusions about the relative power of interaction analysis depend on subjective definitions of comparable effects, is it possible to draw any general conclusions? The argument here is that there are, in fact, definitions which could be acceptable for many situations. For example, in the context of the 2x2 ANOVA design, it seems very reasonable that, since the simple main and main effects are similar in structure, equal threshold effect sizes should be specified. Furthermore, it seems reasonable to assume that, if a certain threshold value has been specified for a pairwise difference within a row or column, it would be desirable to detect if the difference changes by the same amount across the two levels of the other factor. That is, it would be reasonable to specify the same threshold value for the "difference of differences" interaction as that used for the simple main and main effects under the difference definition.

The reasoning would be similar for the ATI case. That is, a threshold value for the aptitude main effect would be set. It would then be desirable if aptitude simple main effects and a difference in the aptitude simple main effects (i.e., the slopes for the two treatments and the difference in the two slopes) of the same magnitude could be detected. This reasoning is similar to that of Cronbach and Snow (1977, p. 82) which was cited earlier.

The above reasoning for both the ANOVA and ATI models can be summarized by saying that an appropriate definition of comparable effects corresponds to assuming equal effects under the difference definition (or, alternatively, to assuming $(\alpha_i)_j/\alpha_i = 1$ and $\gamma_{ij}/\alpha_i = 1/2$ under the deviation definition). If this reasoning is accepted, the relative weakness of ATI analysis discussed by Cronbach and Snow (1977) is also present in interaction analysis in general. That is, based on the power comparisons in Table 1 which apply to both the factorial ANOVA and the ATI models, the power for the simple main effects considered here would be .63, and that for the interaction would be .37, compared to a power of .90 for the main effect. Furthermore, under this same reasoning for models with more independent variables (i.e., for higher order factorial ANOVA designs or ATI models with more than one aptitude or treatment variable), the power for higher order simple and interaction effects drops sharply.

REFERENCES

Cohen, J. _Statistical power analysis for the behavioral sciences_ (Revised edition). New York: Academic Press, 1977.

Cronbach, L. & Snow, R. _Aptitudes and instructional methods_. New York: Irvington Publishers, 1977.