Chapter 7

Effect Size Estimation in Multifactor Designs

I believe that a scientist looking at nonscientific problems is just as dumb as the next guy.

—Richard Feynman

Overview

- Special considerations
- Types of multiple-factor designs
- Factorial ANOVA
- Analysis options for nonorthogonal designs
- Standardized contrasts
- Measures of association

Special considerations

- Morris and DeShon (1997) and others note there are some special considerations in designs with multiple factors
- Ignoring these special issues can result in effect size variation across multiple-factor designs due more to artifacts than real differences in effect size magnitude
- These issues arise in part out of the problem of what to do about other factors when effects on a particular factor are estimated
- Methods for effect size estimation in multiple-factor designs are also not as well developed as for single-factor designs
- However, this is more true for standardized contrasts than for measures of association

Special considerations

- It is generally necessary to have a good understanding of ANOVA for multiple-factor designs (e.g., factorial ANOVA) in order to understand effect size estimation

- However, this does not mean one needs to know about the *F* test only, which is just a small (and perhaps the least interesting) part of ANOVA

- The most useful part of an ANOVA source table for the sake of effect size estimation is everything to the left of the usual columns for *F* and *p*

- It is better to see ANOVA as a general tool for estimating variance components for different types of effects in different kinds of designs

Types of multiple-factor designs

- Multiple-factor designs arise out of a few basic distinctions, including whether the

   1. factors are between-subjects vs. within-subjects
   2. factors are experimental vs. nonexperimental
   3. relation between the factors or subjects is crossed vs. nested

- The most common type of multiple-factor design is the *factorial design*, in which every pair of factors is *crossed* (levels of each factor are studied in all combinations with levels of all other factors)

Types of multiple-factor designs

- Some common types of factorial designs:
  - *Completely between-subjects factorial design*:

    Subjects nested under all combinations of factor levels (i.e., all samples are independent)
  - *Mixed within-subjects factorial design* (*split-plot* or *mixed design*):

    At least one factor is between-subjects and another is within-subjects (i.e., subjects are crossed with some factors but nested under others)
  - *Completely within-subjects factorial design*:

    Each case in a single sample is tested under every combination of two or more factors (i.e., subjects are crossed with all factors)

Types of multiple-factor designs

- Variations on factorial designs (not as common):
  - *Hierarchical design*:

    At least one factor is nested under another factor, such as a control factor that is considered random

  - *Partial or incomplete designs*:

    Levels of some factors are not studied in every combination with levels of other factors, which reduces the number of cells (i.e., fewer cases are needed)

Factorial ANOVA

- Factorial ANOVA also arises out of a few basic distinctions, including whether the

    1. factors are between-subjects vs. within-subjects

    2. factors are fixed vs. random

    3. cell sizes are equal vs. unequal

Factorial ANOVA

- Not all these distinctions are necessary—for example:
  - Random models and repeated-measures models are related
  - Repeated measures can be analyzed with basic computer programs for factorial ANOVA without special capabilities for handing repeated observations (see Schuster & von Eye, 2001)
- The first two distinctions affect the denominator (error term) of the *F* statistic, but not the derivation of effect sum of squares (see Frederick, 1999)

Factorial ANOVA

- A distinction that cannot be ignored is whether a factorial design is balanced or not
- In a balanced (equal-n) design, the main and interaction effects are all independent
- For this reason balanced factorials are referred to as *orthogonal designs*

Factorial ANOVA

- However, factorial designs in applied research are often not balanced (e.g., Keselman et al., 1998)

- We must distinguish between unbalanced designs with

    1. unequal but proportional cell sizes

    2. unequal and disproportional cell sizes

- This is because designs with proportional cell sizes can be analyzed as orthogonal designs (e.g., Keren, 1993)

Factorial ANOVA

- In *nonorthogonal designs*, the main effects overlap (i.e., they are not independent)
- This overlap can be corrected in different ways, which means that there may be no unique estimate of the sums of squares for a particular main effect
- This ambiguity can affect both statistical tests and effect size estimates, especially measures of association
- The choice among alternative sets of estimates for a particular nonorthogonal design is best based on rational considerations, not statistical ones

Factorial ANOVA

- Just as in single-factor ANOVA, two basic sources of variability are estimated in factorial ANOVA, within-conditions and between-conditions
- The total within-conditions variance, MS_W, is estimated the same basic way—as the weighted average of the within-conditions variances (e.g., Equation 7.1 for a two-factor design)
- This is true regardless of whether the design is balanced or not

Factorial ANOVA

- However, estimation of the numerator of the total between-conditions variability in a factorial design depends on whether the design is balanced or not
- The “standard” equations for effect sums of squares (see Table 7.2) presented in many introductory statistics books are for balanced designs only
- It is only for such designs that the sums of squares for the main and interaction effects are both additive and unique

Factorial ANOVA

- For example, it is only in an orthogonal two-way design that the following relations always hold:

$$SS_{A,\ B,\ AB} = SS_A + SS_B + SS_{AB}$$

$$\hat{\eta}^2_{A,\ B,\ AB} = \hat{\eta}^2_A + \hat{\eta}^2_B + \hat{\eta}^2_{AB}$$

- That is, both the total between-conditions sums of squares and the total proportion of explained variance can be uniquely decomposed into values for the individual main and interaction effects

Factorial ANOVA

- Effect definitions (e.g., see Table 7.1 for a two-way design)
- Single-factor effects:
  - *Main effect* :

    Concerns variability of the marginal means for each factor around the grand mean—estimated by collapsing across the levels of the other factor

  - *Simple effect* :

    Concerns variability of cell means in the same row or column—estimated controlling for the one level of the other factor

Factorial ANOVA

- *Interaction effects* can be seen in a few different ways:
  - A conditional effect where the simple effects of each factor on the outcome variable are different across the levels of the other factor
  - Also known as *moderator effects*: The impact of one factor depends on another factor
  - A *mediator effect* is something different: An indirect effect of the type estimated in path analysis

Factorial ANOVA

- Contrasts can be specified for main, simple, or interaction effects where *df* > 1 (i.e., they are omnibus)

- A single-factor contrast involves the levels of just one factor while we are controlling for the other factors—there are two kinds:

    1. A *main comparison* involves contrasts between subsets of marginal means for the same factor (i.e., it is conducted within a main effect)

    2. A *simple comparison* involves contrasts between subsets of cell means in the same row or column (i.e., it is conducted within a simple effect)

- A single-factor contrast is specified with weights just as a contrast in a one-way design—we assume here mean difference scaling (i.e., $\sum | c_i | = 2.0$)

Factorial ANOVA

- An interaction contrast specifies a single-*df* interaction effect
- It is specified with weights applied to cells of the whole design (e.g., to all six cells of a 2×3 design)
- These weights follow the same general rules as for one-way designs

Factorial ANOVA

- The weights should also be d*oubly centered*, which means that they sum to zero in any row or column
- If an interaction contrast in a two-way design should be interpreted as the difference between a pair of simple comparisons (i.e., mean difference scaling), the sum of the absolute values of the weights must be 4.0 (see Bird, 2002)

Factorial ANOVA

- Example of weights for a simple interaction contrast:

	B_1	B_2	B_3
A_1	1	0	−1
A_2	−1	0	1

- Note that these weights are doubly centered and are a standard set for a two-way design because $\sum | c_i | = 4.0$
- These weights compare the simple effect of A at B_1 with the simple effect of A at B_3

Factorial ANOVA

- It should be obvious by now that there can be many possible effects that can be analyzed in a factorial design
- This is especially true for designs with three or more factors, for which there are a three-way interaction effect, two-way interaction effects, main effects, and contrasts for any of the effects just listed where $df > 1$
- One can easily get lost by estimating every possible effect

Factorial ANOVA

- This means it is *critical* to have a plan that minimizes the number of analyses while still respecting essential hypotheses
- Some of the worst misuses of statistical tests are seen in factorial designs when this advice is ignored—example:

> All possible effects are tested and sorted into two categories, those statistically significant and subsequently discussed at length vs. those not statistically significant and subsequently ignored

- This misuse is compounded when power is ignored, which can vary from effect to effect in factorial designs

Analysis options for nonorthogonal designs

- This is a brief review of options for estimating effect sum of squares in nonorthogonal factorial designs—see Keren (1993) and Rencher (1998) for more information
- Different methods correct for overlap of main effects in different ways
- For this reason they do not always generate the same set of sums of squares values for a particular data set
- With no rationale to select among different sets of estimates, they should all be seen as equally correct
- Most analysis options for unbalanced designs in contemporary computer programs for factorial ANOVA are based on regression methods

Analysis options for nonorthogonal designs

- Although not exhaustive, this typology by Overall and Spiegel (1969) is helpful:
  - Method 1 estimates effect sums of squares controlling for all other effects—these sums of squares may be labeled “Type III” or “unique” in program output
  - Method 2 adjusts sums of squares for the main effects for overlap with each other—these sums of squares may be labeled “Type II” or “classical experimental” in program output
  - Method 3 does not remove shared variance from the sums of squares of one main effect (e.g., *A*) but adjusts the sums of squares of the other main effect for overlap with the first (e.g., *B* adjusted for A)—these sums of squares may be labeled “Type I,” “sequential,” or “hierarchical” in program output

Analysis options for nonorthogonal designs

- Some suggestions:
  - Method 1 does not generally give greater weight to cells with more observations, so this method may be optimal when unequal cell sizes result from random data loss from a few cells
  - Method 2 and Method 3 may be a better choice for nonexperimental designs where unequal cell sizes reflect unequal group sizes in the population—this is because they permit the actual cell sizes to contribute to the analysis but with different priorities given to certain main effects
- Examples of specific rationales for selecting one method over another for unbalanced designs with nonexperimental factors are given in the research examples for this chapter

Standardized contrasts

- There is no definitive method at present for calculating standardized contrasts in factorial designs
- However, some general principles discussed by Cortina and Nouri (2000), Olejnik and Algina (2000), and others are that:

  1. Estimates for effects of each independent variable in a factorial design should be comparable with effect sizes for the same factor studied in a one-way design

  2. Changing the number of factors in the design should not necessarily change the effect size estimates for any one of them

Standardized contrasts

- Standardized contrasts may be preferred over measures of association as effect size indexes if contrasts are the main focus of the analysis
- This is most likely in designs with just two factors
- Because they are more efficient, measures of association may be preferred in larger designs—for example, the total predictive power in a two-way design is summarized by $\hat{\eta}^2_{A,\ B,\ AB}$

Standardized contrasts

- Standardized contrasts in factorial designs have the same general form as in one-way designs: $d_{\hat{\psi}} = \hat{\psi}/\hat{\sigma}^*$
- The problem is figuring out which standardizer should go in the denominator of $d_{\hat{\psi}}$
- The approach described in the text is for single-factor contrasts, such as main comparisons or simple comparisons

Standardized contrasts

- In a two-way design, this approach distinguishes the *factor of interest* and the *off-factor*
- Suppose in a two-way design that two levels of factor A are compared at only one level of factor B (e.g., A_1 vs. A_2 at B_2)
- The factor of interest is A, and the off-factor is B

Standardized contrasts

- If the off-factor *B* does *not* vary naturally in the population to which the results should generalize, it is fine that the standardizer should control for factor *B*

- In this case the square root of MS_W could be the standardizer (i.e., $d_{\hat{\psi}}$ is calculated as $g_{\hat{\psi}}$ in a one-way design)

- If the off-factor *B* does vary naturally, however, the standardizer should *not* control for its variability

- In this case a better standardizer would be the square root of $MS_{W,\, B,\, AB}$, the adjusted within-conditions variance where variability due to factor *B* is pooled with the within-cells variability

Standardized contrasts

- Different methods to calculate $MS_{W,\ B,\ AB}$ are discussed in the text, including:

  1. The *orthogonal sums of squares method* (Glass, McGaw, & Smith, 1981), which requires additive sums of squares (e.g., from Method 3 for an unbalanced design; see Table 7.6)

  2. The *reduced cross-classification method* (Olejnik & Algina, 2000), which works just as well in unbalanced or unbalanced designs

Standardized contrasts

- Only the first method (orthogonal sums of squares) is demonstrated here: $MS_{W,\,B,\,AB}$ would be estimated as follows:

$$MS_{W,\,B,\,AB} = \frac{SS_W + SS_B + SS_{AB}}{df_W + df_B + df_{AB}} = \frac{SS_T - SS_A}{df_T - df_A}$$

- This adjusted variance equals MS_W in the one-way ANOVA with just factor *A*

- It does *not* underestimate the population variance when *B* varies naturally, but MS_W from the two-way ANOVA does so

Standardized contrasts

- Note there is little statistical research to support the method just described
- It can also be difficult to apply to interaction contrasts
- Many additional computational examples can be found in Cortina and Nouri (2000) and Olejnik and Algina (2000)

Standardized contrasts

- At present there is relatively little software support for calculating standardized contrasts in factorial designs much less confidence intervals for them
- An exception is the PSY program, which analyzes raw data from factorial designs with one or more between-subjects factors and/or one or more within-subjects factors and calculates approximate confidence intervals for δ_ψ
- However, the program does not distinguish among multiple between-subjects factors or multiple within-subjects factors
- This means that PSY in its current form is easiest to use in mixed factorial designs with one between-subjects factor and one within-subjects factor (see Bird, 2002)

Measures of association

- If one understands measures of association for one-way designs, there is relatively new to learn about use of these indexes for factorial designs

- The descriptive statistics $\hat{\eta}^2$ and partial $\hat{\eta}^2$ for designs with fixed factors have the same general form in one-way and factorial designs

- It is fairly common in a factorial design to calculate $\hat{\eta}^2$ for all effects together and partial $\hat{\eta}^2$ for individual effects—for example:

    In a two-way design, calculate $\hat{\eta}^2_{A,\,B,\,AB}$ for the total effects and partial $\hat{\eta}^2_A$, partial $\hat{\eta}^2_B$, and partial $\hat{\eta}^2_{AB}$ for the individual effects

Measures of association

- The inferential statistics $\hat{\omega}^2$ and partial $\hat{\omega}^2$ for balanced designs with fixed factors and the intraclass correlation $\hat{\rho}_I$ for balanced designs with random factors also have the same general form in one-way and factorial designs

- See Tables for 7.7 and 7.8 for equations for variance component estimators for $\hat{\omega}^2$ and $\hat{\rho}_I$ for some common types of two-way factorial designs

- Equations for additional types of factorials designs can be found in Dodd and Schultz (1973) and Vaughn and Corballis (1969)

- Some of the same computer programs or macros described in the previous chapter for one-way designs can also be used to construct confidence intervals based on measures of association in some kinds of factorial designs

Measures of association

- Examples:
  - A macro for SPSS by Smithson (2001) calculates noncentral confidence intervals based on $\hat{\eta}^2$ for the total effects in a completely between-subjects design with fixed factors and confidence intervals based on partial $\hat{\eta}^2$ for all other effects (see slides for chap. 4 for the Web addresses)
  - The Power Analysis module in STATISTICA can also calculate noncentral confidence intervals based on $\hat{\eta}^2$ in similar kinds of designs (see Steiger & Fouladi, 1997)
  - Fidler and Thompson (2001) list SPSS macros for calculating noncentral confidence intervals based on $\hat{\omega}^2$ and partial $\hat{\omega}^2$ in balanced, completely between-subjects design with fixed factors (see also Burdick & Graybill, 1992)

References

Bird, K. D. (2002). Confidence intervals for effect sizes in analysis of variance. *Educational and Psychological Measurement*, *62*, 197-226.

Burdick, R. K., & Graybill, F. A. (1992). *Confidence intervals for variance components*. New York: Marcel Dekker.

Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, CA: Sage.

Dodd, D. H., & Schultz, R. F. (1973). Computational procedures for estimating magnitude of effect for some analysis of variance designs. *Psychological Bulletin*, *79*, 391-395.

Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random-effects effect sizes. *Educational and Psychological Measurement*, *61*, 575-604.

Frederick, B. N. (1999). Fixed-, random-, and mixed-effects ANOVA models: A user-friendly guide for increasing the generalizability of ANOVA results. In B. Thompson (Ed.), *Advanced in social science methodology* (Vol. 5, pp. 111-122). Stamford, CT: JAI Press.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage.

Keren, G. (1993). A balanced approach to unbalanced designs. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Vol. 2. Statistical issues* (pp. 155-186). Hillsdale, NJ: Erlbaum.

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., et al. (1998). Statistical practices of education researchers: An analysis of the ANOVA, MANOVA, and ANCOVA analyses. *Review of Educational Research*, *68*, 350-386.

Morris, S. B., & DeShon, R. P. (1997). Correcting effect sizes computed with factorial analyses of variance for use in meta-analysis. *Psychological Methods*, *2*, 192-199.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational Psychology*, *25*, 241-286.

Overall, J. E., & Spiegel, D. K. (1969). Concerning least-squares analysis of experimental data. *Psychological Bulletin*, *72*, 311-322.

Rencher, A. C. (1998). *Methods of multivariate analysis: Vol. 2. Multivariate statistical inference and applications*. New York: Wiley.

Schuster, C., & von Eye, A. (2001). The relationship of ANOVA models with random effects and repeated measurement designs. *Journal of Adolescent Research*, *16*, 205-220.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement*, *61*, 605-632.

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Erlbaum.

Vaughn, G. M., & Corballis, M. C. (1969). Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. *Psychological Bulletin*, *72*, 204-213.