FACTORIAL EFFECT SIZE

Between-subjects design

Multiple factors

- 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*
  - The sums of squares and mean squares
- As we have mentioned, it is better to see ANOVA as a general tool for estimating variance components for different types of effects in different kinds of designs

Multiple factors

- 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)

Multiple factors

- 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)
- *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)
- *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)

Multiple factors

- While there are other distinctions, one 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*

Multiple factors

- However, factorial designs in applied research are often not balanced
- We have distinguished between unbalanced designs with
  - 1. unequal but proportional cell sizes
  - 2. unequal and disproportional cell sizes
- As we noted designs with proportional cell sizes can be analyzed as orthogonal designs as equal n is a special case of being proportional

Multiple factors

- 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
  - If for example two main effects are correlated, how would you know if factor B is significant or simply significant because of its correlation with A? Its effect could be A
- 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-subjects and between-subjects
- The total within-conditions variance, MS_W, is estimated the same basic way—as the weighted average of the within-conditions variances
  - This is true regardless of whether the design is balanced or not
- 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 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
  - i.e. we could add up the eta^2 for main effects and interaction to get the overall between subjects eta^2

Factorial ANOVA

- As we have noted, contrasts can be specified for main, simple, or interaction effects
- 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., $\sum |a_i| = 2.0$)

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, but see your Kline text for specifics regarding effect size estimation
  - 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
- As an example, these weights compare the simple effect of *A* at *B*1 with the simple effect of *A* at *B*3

	B_1	B_2	B_3
A_1	1	0	−1
A_2	−1	0	1

Factorial ANOVA

- As is probably clear 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
- One can easily get lost by estimating every possible effect which means it is important 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
  - E.g. 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

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 from study to study 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 indices 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

Standardized Contrasts

- Standardized contrasts in factorial designs have the same general form as in one-way designs: $d = \Psi/\sigma^*$
- The problem is figuring out which standardizer should serve as the denominator
- Here we'll present what is described in both Kline and Howell in a general way, though Kline has more specifics

Standardized Contrasts

- Basically it comes down to putting the variance due to other effects not being considered back into the error variance which will become our standardizer
- So if in our ANOVA we have SS for
    - Therapy
    - Gender
    - Therapy*Gemder
    - Err
  - If we used the $SS_{err} \rightarrow MS_{err} \rightarrow \sqrt{MS_{err}}$ here as we have elsewhere for a standardizer, it would be much smaller than in a one way design comparing groups in main effect A only (for example)
- Then we'd run around saying we had a much larger effect than those who just looked at a main effect of therapy, everyone

Standardized Contrasts

- The solution then is to add those other effects back into the error term

$$\hat{s} = \sqrt{\frac{SS_{gender} + SS_{gender\ X\ therapy} + SS_{error}}{df_{gender} + df_{therapy} + df_{error}}}$$

- What of simple effects?
- The simple effects would use the same standardizer as would be appropriate for the corresponding main effect

Measures of Association

- If using measures of association, the process is the same as in the one-way design

$$\omega^2 = \frac{SS_{effect} - (k-1)MS_{error}}{SS_{total} + MS_{error}}$$

Summary

- As with the one-way design we have the approach of looking at standardized mean differences (d-family) or variance accounted for assessments of effect size (r-family)
- There is no change as far as the r-family goes when dealing with factorial design
- The goal for standardized contrasts is to come up with a measure that will reflect what is seen for main effects *and* be consistent across studies of different types