Chapter 6

Effect Size Estimation in One-Way Designs

To live a creative life, we must
lose our fear of being wrong.

—Joseph Chilton Pearce

Overview

- Contrast specification and tests
- Standardized contrasts
- Measures of association
- Effect size estimation in covariate analyses

Contrast specification and tests

- Concerns design with a single factor *A* with $a \geq 2$ levels (conditions)
- The *omnibus comparison* concerns all levels (i.e., $df_A \geq 2$)
- A *focused comparison* or *contrast* concerns just two levels (i.e., $df = 1$)
- The omnibus effect is often relatively uninteresting compared with specific contrasts (e.g., treatment 1 vs. placebo control)
- A large omnibus effect can also be misleading if due to a single discrepant mean that is not of substantive interest

Contrast specification and tests

- Traditional approach is to analyze the omnibus effect followed by analysis of all possible pairwise contrasts (i.e., compare each condition to every other condition)
- However, this approach is typically incorrect (Wilkinson & TFSI, 1999)—for example, it is rare that all such contrasts are interesting
- Also, use of traditional methods for post hoc comparisons (e.g., Newman-Keuls) reduces power for every contrast, and power may already be low

Contrast specification and tests

- A contrast is
  - a directional effect that corresponds to a particular facet of the omnibus effect
  - often represented with the symbol ψ for a population or $\hat{\psi}$ for a sample
  - a weighted sum of means
- In a sample, a contrast is calculated as:

$$\hat{\psi} = \sum_{i=1}^{a} c_i \, M_i$$

$c_1, c_2, \ldots, c_a$ is the set of weights that specifies the contrast

Contrast specification and tests

- Contrast weights must sum to zero and weights for at least two different means should not equal zero

- Means assigned a weight of zero are excluded from the contrast

- Means with positive weights are compared with means given negative weights (e.g., see Table 6.1)

Contrast specification and tests

- For effect size estimation with the *d* family, we generally want a *standard set* of contrast weights, also referred to as *mean difference scaling* (Bird, 2002; Keppel, 1991)
- In a one-way design, the sum of the absolute values of the weights in a standard set equals two (i.e., $\sum | c_i | = 2.0$)
- Mean difference scaling permits the interpretation of a contrast as the difference between the averages of two subsets of means

Contrast specification and tests

- An exception to the need for mean difference scaling is for *trends* (*polynomials*) specified for a quantitative factor (e.g., drug dosage)
- There are default sets of weights that define trend components (e.g., linear, quadratic, etc.) that are not typically based on mean difference scaling
- Not usually a problem because effect size for trends is generally estimated with the *r* family (measures of association)
- Measures of association for contrasts of any kind generally correct for the scale of the contrast weights

Contrast specification and tests

- Two contrasts are orthogonal if they each reflect an independent aspect of the omnibus effect

- In balanced designs, two contrasts are orthogonal if the following equality holds:

$$\sum_{i=1}^{a} c_{1_i} c_{2_i} = 0$$

- In unbalanced designs, two contrasts are orthogonal if the following equality holds:

$$\sum_{i=1}^{a} \frac{c_{1_i} c_{2_i}}{n_i} = 0$$

Contrast specification and tests

- The maximum number of orthogonal contrasts is $df_A = a - 1$
- For a set of all possible orthogonal pairwise contrasts:

$$SS_A = \sum_{i=1}^{a-1} SS_{\hat{\psi}_i} \qquad \text{and} \qquad \hat{\eta}_A^2 = \sum_{i=1}^{a-1} \hat{\eta}_{\hat{\psi}_i}^2$$

- That is, the omnibus effect can be broken down into a – 1 independent directional effects
- However, it is more important to analyze contrasts of substantive interest even if they are not orthogonal

Contrast specification and tests

- The *t* statistic for a contrast between independent means for a nil hypothesis is:

$$t_{\hat{\psi}}\ (df_W) = \frac{\hat{\psi}}{s_{\hat{\psi}}}$$

$$s_{\hat{\psi}} = \sqrt{MS_W \left(\sum_{i=1}^{a} \frac{c_i^2}{n_i} \right)}$$

Contrast specification and tests

- The *F* statistic for a contrast between independent means is:

$$F_{\hat{\psi}}\,(1,\, df_W) \;=\; t^2_{\hat{\psi}}\,(df_W) \;=\; \frac{SS_{\hat{\psi}}}{MS_W}$$

$$SS_{\hat{\psi}} \;=\; \frac{\hat{\psi}^2}{\sum_{i=1}^{a} \frac{c_i^2}{n_i}}$$

- Both test statistics for contrasts ($t_{\hat{\psi}}, F_{\hat{\psi}}$) assume homogeneity of variance

Contrast specification and tests

- Test statistics for dependent mean contrasts usually have error terms based on only the two conditions compared—for example:

$$t_{\hat{\psi}}\ (n - 1) = \frac{\hat{\psi}}{s_{\hat{\psi}}}$$

$$s_{\hat{\psi}} = \sqrt{\frac{s^2_{D_{\hat{\psi}}}}{n}}$$

$s^2_{D_{\hat{\psi}}}$ is the variance of the contrast difference scores ($D_{\hat{\psi}}$)

- This is because such error terms do not assume sphericity, and the sphericity assumption is often violated in behavioral data

Contrast specification and tests

- Distributions of contrasts are simple, so traditional confidence intervals based on them are generally fine
- The general form of an individual confidence interval for ψ is:

$$\hat{\psi} \pm s_{\hat{\psi}} \left[t_{\hat{\psi}_{2\text{-tail},\, \alpha}} \left(df_{\text{error}} \right) \right]$$

df_{error} is the degrees of freedom from the ANOVA error for the contrast

Contrast specification and tests

- There are also corrected confidence intervals for contrasts that adjust for multiple comparisons (i.e., inflated Type I error)

- They are known as *simultaneous* or *joint confidence intervals*

- Their widths are generally narrower compared with individual confidence intervals because they are based on a more conservative value of $t_{\hat{\psi}_{\text{2-tail},\,\alpha}}$

- The freely available program PSY can automatically calculate simultaneous confidence intervals (see also Bird, 2002):

  http://www.psy.unsw.edu.au/research/PSY.htm

Standardized contrasts

- The parameter for a standardized contrast is:

$$\delta_{\psi} = \psi / \sigma^*$$

σ^* is a population standard deviation with more than one form

- The general form of a sample standardized contrast is:

$$d_{\hat{\psi}} = \hat{\psi} / \hat{\sigma}^*$$

Standardized contrasts

- There are three general ways to estimate σ* (i.e., the standardizer) for contrasts between independent means:

    1. Calculate $d_{\hat{\psi}}$ as Glass's Δ (i.e., use the standard deviation of the control group)

    2. Calculate $d_{\hat{\psi}}$ as Hedge's *g* (i.e., use the square root of the pooled within-conditions variance for just the two groups being compared)

    3. Calculate $d_{\hat{\psi}}$ as $g_{\hat{\psi}}$ (an extension of *g*) where the standardizer is the square root of MS_W based on all groups (generally recommended)

Standardized contrasts

- The statistic $g_{\hat{\psi}}$ can also be calculated from $t_{\hat{\psi}}$ for a contrast based on a standard set of weights:

$$g_{\hat{\psi}} = t_{\hat{\psi}} \sqrt{\left(\sum_{i=1}^{a} \frac{c_i^2}{n_i} \right)}$$

Standardized contrasts

- There are two general ways to estimate σ* for a dependent mean change:

    1. Use any method for contrasts between independent means except the above equation, which requires an independent-samples test statistic

    2. Standardize the dependent mean change against the standard deviation of the contrast difference scores, $s_{D_{\hat{\psi}}}$

Standardized contrasts

- The first method just mentioned makes a standardized mean change from a correlated design more directly comparable with a standardized contrast from a design with unrelated samples

- The PSY program standardizes all contrasts against the square root of MS_W (i.e., it calculates $g_{\hat{\psi}}$ for any design it analyzes)

Standardized contrasts

- Distributions of standardized contrasts are generally not simple
- An approximate confidence interval for δ_ψ based on $g_{\hat{\psi}}$ has the general form:

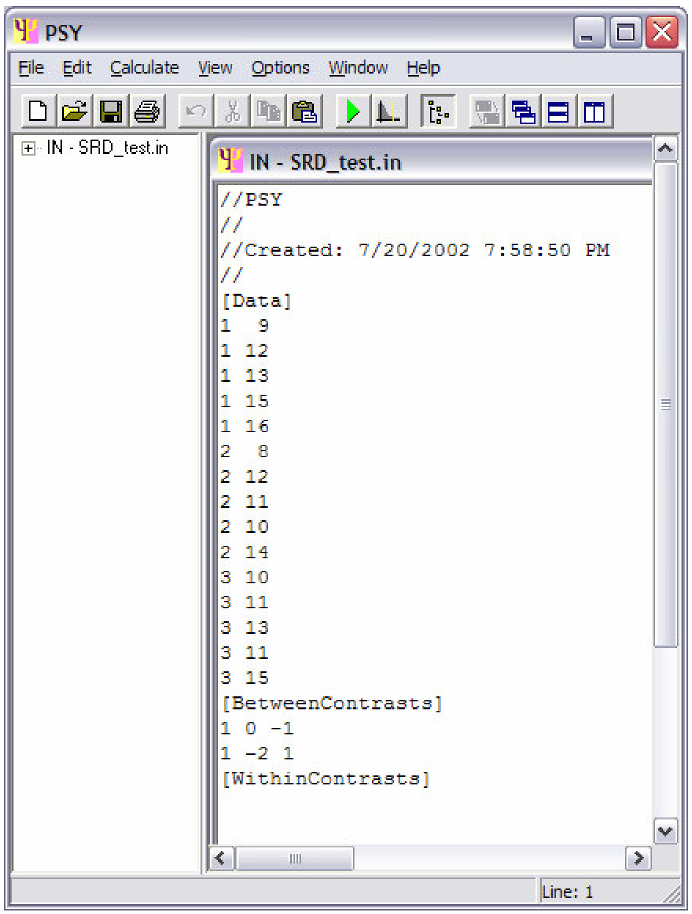
$$g_{\hat{\psi}} \pm s_{g_{\hat{\psi}}} \left[t_{\hat{\psi}_{2\text{-tail}, \alpha}} (df_W) \right]$$

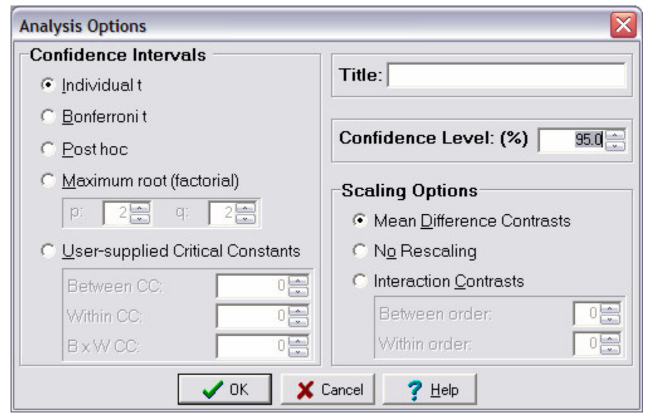
$$s_{g_{\hat{\psi}}} = \sqrt{\sum_{i=1}^{a} \frac{c_i^2}{n_i}}$$

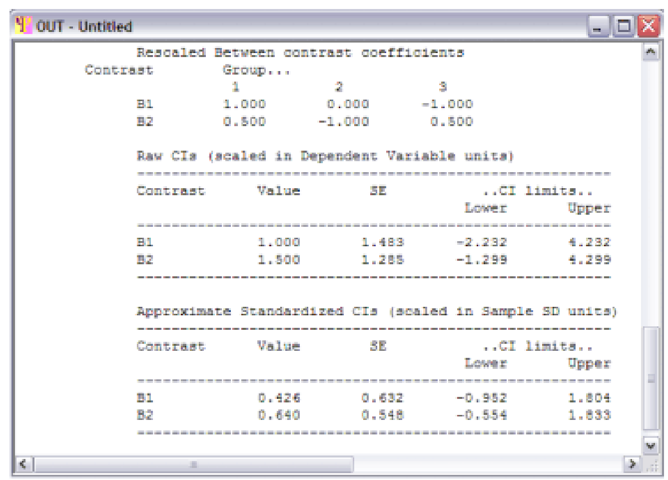
- Bird (2002) recommends this method for designs with either independent or dependent samples whenever the effect size is $g_{\hat{\psi}}$
- An alternative for correlated designs is to standardize the bounds of the confidence interval for ψ based on the standard deviation of the difference scores (i.e., the metric is not that of the original scores)

Standardized contrasts

- Screenshots from use of PSY to calculate $g_{\hat{\psi}}$ and approximate 95% confidence intervals for the data in Table 6.3 (see also Table 6.4):







Standardized contrasts

- Noncentral confidence intervals for δ_ψ are generally more exact than approximate confidence intervals
- They are constructed according to the same basic principles as noncentral confidence intervals for δ in two-sample designs
- A computer tool is required
- For example, the Power Analysis module of STATISTICA or the ESCI program can be used to obtain exact confidence intervals for δ_ψ based on $g_{\hat{\psi}}$ when the samples are independent
- See also code for SAS/STAT in Table 6.6

Measures of association

- For designs with a fixed factor, the correlation ratio can be calculated for any effect:

$$\hat{\eta}^2_{\text{effect}} = \frac{SS_{\text{effect}}}{SS_T}$$

- When the samples are independent:

$$\hat{\eta}^2_A = \frac{F_A}{F_A + df_W / df_A}$$

Measures of association

- It is a fairly common practice to calculate $\hat{\eta}^2_A$ for the omnibus effect but to calculate the partial correlation ratio for each contrast

- General form of the partial correlation ratio:

$$\text{partial } \hat{\eta}^2_{\text{effect}} = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}}$$

- Partial $\hat{\eta}^2_{\text{effect}}$ corrects for all other sources of between-conditions variability other than the effect of interest

- Partial $\hat{\eta}^2_{\text{effect}}$ in a correlated design also corrects for the subjects effect

Measures of association

- The inferential statistics $\hat{\omega}^2_{\text{effect}}$ and partial $\hat{\omega}^2_{\text{effect}}$ adjust for bias in estimation of the population proportion of explained variance
- Both assume fixed factors and a balanced design
- The correlation ratio does not assume a balanced design
- Their values are generally lower than those of the corresponding correlation ratios for the same data, but their values converge as the sample size increases
- Note that the values of $\hat{\omega}^2_{\text{effect}}$ or partial $\hat{\omega}^2_{\text{effect}}$ can be negative—if so, interpret as though the value were zero

Measures of association

- The inferential statistic $\hat{\rho}_I$ is the *intraclass correlation* for comparative designs with random factors—it is already in a squared metric
- The statistic $\hat{\rho}_I$ in a one-way design is ordinarily calculated only for the omnibus effect
- It also assumes a balanced design, and its calculated value can be negative

Measures of association

- The calculation of $\hat{\omega}^2_{\text{effect}}$ and $\hat{\rho}_{\text{I}}$ are based on the estimation of variance components

- A *variance component* corresponds to an element in the ANOVA structural model for a particular design, and its estimate reflects the distributional assumptions for that design

- For example, the distributional model for a random factor is a variation on that for a repeated-measures factor

Measures of association

- The general form of both $\hat{\omega}^2$ and $\hat{\rho}_I$ for the omnibus effect is $\hat{\sigma}^2_\alpha / \hat{\sigma}^2_{tot}$
- The numerator, $\hat{\sigma}^2_\alpha$, estimates population variability at the level of means due to factor *A*
- The denominator, $\hat{\sigma}^2_{tot}$, estimates total variability at the case level for all populations

Measures of association

- The general form of partial $\hat{\omega}^2_{\text{effect}}$ is $\hat{\sigma}^2_{\text{effect}} / (\hat{\sigma}^2_{\text{effect}} + \hat{\sigma}^2_{\varepsilon})$ where $\hat{\sigma}^2_{\varepsilon}$ estimates population error variance

- Specific equations for variance components, such as $\hat{\sigma}^2_{\alpha}$, $\hat{\sigma}^2_{\psi}$, or $\hat{\sigma}^2_{\varepsilon}$, depend on whether the factor is fixed vs. random or between-subjects vs. within-subjects

- The formulas in Tables 6.7-6.9 are based on sources such as Dodd and Schultz (1973), Kirk (1995), Myers and Well (2002), Vaughn and Corballis (1969), and Winer, Brown, and Michels (1991)

- They are based on an ANOVA-approach to variance component estimation, but there are other methods (e.g., maximum likelihood)

Effect size estimation in covariate analyses

- This discussion deals only with between-subjects designs with a single fixed factor and one covariate
- However, the principles generalize to other kinds of covariate analyses
- It also assumes standardization based on a pooled within-groups standard deviation

Effect size estimation in covariate analyses

- It should be noted that the statistical assumptions of the analysis of covariance (ANCOVA) are *very* restrictive—they include:

    1. Usual assumptions for ANOVA (e.g., normality, homogeneity of variance)

    2. Covariate scores are perfectly reliable

    3. Homogeneity of regression of the outcome variable on the covariate across all populations

Effect size estimation in covariate analyses

- ANCOVA also yields estimated group means on the outcome variable adjusted for the covariate
- Correct interpretation of these adjusted means (and other ANCOVA results) may require the assumption that the covariate is unrelated to the independent variable
- This assumption is usually tenable in experimental designs, but often not in nonexperimental designs
- If the covariate is related to the independent variable but is not the only variable in a nonexperimental design on which nonequivalent groups differ, ANCOVA results may not be accurate
- See Campbell and Erlebacher (1975), Cook and Campbell (1979, chap. 3)

Effect size estimation in covariate analyses

- There are two choices for the numerator of a standardized contrast in ANCOVA: the unadjusted mean contrast or the mean contrast adjusted for the covariate

- There should be little difference between these two values in an experimental design

- There are two choices for the denominator:

    1. The square root of MS_W, which is not adjusted for the covariate (i.e., the ANOVA error term)

    2. The adjusted error term in the ANCOVA

Effect size estimation in covariate analyses

- Cortina and Nouri (2000) suggest that if the covariate varies naturally in the population to which the results should generalize, the square root of MS_W may be the best choice as the standardizer

- In general, calculate a standardized contrast in ANCOVA as $g_{\hat{\psi}}$ (i.e., just as in the ANOVA for the same data)

Effect size estimation in covariate analyses

- Measures of association in ANCOVA can also be calculated from different perspectives (i.e., adjusting for the covariate or not)
- The correlation ratio:

$$\hat{\eta}_A^{2\prime} = \frac{SS_A^{\prime}}{SS_T^{\prime}}$$

  where both sums of squares terms are adjusted for the covariate is actually the difference between two squared multiple correlations

- One of these squared correlations is with just factor *A* in the equation, and the other is with both factor *A* and the covariate
- The statistic $\hat{\eta}_A^{2\prime}$ estimates the *increase* in total explained variance due to factor *A* above and beyond the predictive power of the covariate (e.g., see Table 6.11)

References

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

Campbell, D. T., & Erlebacher, A. (1975). How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In M. Guttentag & E. L. Struening (Eds.), *Handbook of evaluation research* (Vol. 1, pp. 597-617). Beverly Hills, CA: Sage.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. Chicago: Rand McNally.

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.

Keppel, G. (1991). *Design and analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Kirk, R. E. (1995). *Experimental design* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Myers, J. L., & Well, A. D. (2002). *Research design and statistical analysis* (2nd ed.). 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.

Wilkinson, L., & the Task Force on Statistical Inference (TFSI). (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design* (3rd ed.) Boston: McGraw-Hill.