CHAPTER 2

Types of Effect size indices: An Overview of the Literature

There are different types of effect size indices as a result of their different interpretations. Huberty (2002) names three different types:

- Group differences
- Relationships
- Group overlapping

Before we provide an overview of these types, we will pause a moment to look at the measurement scales and the assumptions, each of which produce different indices.

2.1 Measurement scales and assumptions

The different effect size indices do not only go hand-in-hand with the statistical analyses or objectives to which one strives, but they are also dependent on the type of measurement. For *continuous* measurements (i.e., which can take on any value within a given interval, e.g., a person's height or IQ), sample means are used to compare different populations, while Pearson's correlation coefficient is used to indicate relationships between the different measurements that can be made in the sample. If we are working with *categorical* data (i.e., data whose measurements are divided into categories, e.g., language groups or gender of people), then proportions can be used to make comparisons between populations and measures obtained from two-way frequency tables can be used to determine relationships. In the discussion that follows we will distinguish between effect size indices based on the following measurement *scales*:

- Interval/ratio (continuous measurements, for example, IQ, temperature, blood pressure).

- Ordinal (categorical measurement with ordered categories, e.g., employment position in a company or academic qualification of an individual).
- Nominal (categorical with no ordering, e.g., racial group or church membership of an individual).
- Dichotomous (nominal with only two categories, e.g., someone's HIV-status as positive/negative or gender as Male/Female).

In addition to the type of scale, the effect size indices also depend on the *situation* or *assumptions* underlying the populations or models being used.

The *situation* which one finds oneself in is determined by:

- The number of populations or groups,
- If one is studying the *entire population* (for example, if one wants to only analyse the data from the respondents of a questionnaire and there is no interest in generalizing the results further) or if one is studying only a random *sample,* and
- If *univariate* or *multivariate* effect sizes must be calculated.

The *assumptions* that one might consider include:

- The assumption of *homogeneity of variances* (i.e., the equality of standard deviations) of interval scaled measurements between different populations when effect size indices are based on means or contrasts. Particular attention will be paid to this in the next chapter.
- Underlying models are usually assumed to consist of *fixed factor effects* or *random factor effect* or a combination of the two. Different effect size indices will be considered for each of these assumptions. Chapter 6 will provide further details on these topics.
- When *confidence intervals* for effect size indices are calculated, the assumption of *normality* of the interval scaled measurements is important.

- *Independent* groups (e.g., control and experimental groups) and *dependent* groups (e.g., a group before a treatment, and the same group after a treatment) each have their own specific effect size indices (see Chapter 4).
- Using *covariates for correction* (in analysis of covariance or ANCOVA) introduces its own modifications to effect size indices (see paragraph 6.4).
- When one is concerned with the *reliability* of a variable, then an *attenuation correction* for effect sizes can also be introduced (see paragraph 5.1.4).

2.2 Effect size indices for differences in groups

In a situation where the means of *two* groups must be compared, Cohen (1969, 1977, 1988) proposed the use of the standardized difference of the means (*d*). In the 1970's and 1980's there was a great deal of discussion as to how the difference in the means $(\mu_1 - \mu_2)$ should be standardized. Cohen suggested that one should simply divide it by the common standard deviation (σ) of the two populations, since this assumes homogeneity of variances. Glass (1976) proposed that only the standard deviation of the control group should be used in the denominator. Hedges (1981), in contrast, modified *d* so that it is an unbiased estimator for random samples. For dichotomous variables Cohen (1969, 1977, 1988) compares two proportions with the index *h*, which is the difference between the arc-sine transformations of the proportions. For differences between two means where heterogeneity of variances is assumed, Steyn (2000) proposes an adjusted index Δ_a along with a "conservative" estimator $\hat{\Delta}_a$. These indices could also be applied when comparing proportions (Steyn 1999).

When *more than two groups* are compared, Cohen (1969, 1977, 1988) proposes an index *f* which represents the variance of the group means relative to the common variance σ. He also recommends using $\delta = (\mu_{max} - \mu_{min})/\sigma$ as an index, where μ_{max} and μ_{min} are the largest and smallest population means respectively. A discussion concerning cases where the effect sizes of *contrasts* are important can be found in both Olejnik & Algina (2000) and Kline (2004a: Chapter 6). The former also contemplates cases where means are adjusted for covariates in an ANCOVA design, as for contrasts within individuals.

2.3 Effect size indices based on Relationships

The familiar Pearson product moment coefficient of correlation (r) is a measure of the linear relationship between two *continuous* or *interval scaled variables.* As such, *r* can be used as an effect size index (Cohen, 1977, Chapter 3) to express the strength of the linear relationship.

In cases where one requires the strength of the relationship between *nominal scaled variables,* Cohen (1977, Chapter 7) provides an effect size, based on the Chi-squared test statistic, called *w*. A special case of this effect size is called the ϕ-coefficient and it is used for *dichotomous variables*.

The correlation between a grouping indicator variable (*x*), which only takes on two values (say 1 and 2) and is used to distinguish between two groups, and the interval scale variable (*y*) denotes the relationship between *y* and the group membership (this is also known as a point bi-serial correlation, r_{pb}). This is another sort of interpretation of effect size, because the more the group means change relative to their standard deviations, the larger the value r_{pb} becomes. Cohen (1969, 1977, 1988) also expressed r_{pb} in terms of *d*, the standardized differences in means. The squared value, r_{pb}^2, in turn, represents the proportion of variation of *y* attributed to its population membership, which is a special case of Pearson's eta-squared (η^2). Eta-squared can serve as an effect size whenever more than two population means are compared, in which case it represents the ratio of the between-group (or model) sum of squares and the total sum of squares of a one-way ANOVA model. Hays (1963) made use of the omega-squared index (ω^2) which recently was redefined (Olejnik & Algina, 2000). When the grouping variable can be considered to be a "random factor", the inter-class coefficient of correlation $(\hat{\rho}_I)$ is the recommended measure of this relationship and the square of this quantity, $\hat{\rho}_I^2$, serves as an estimator for η^2 (Olejnik & Algina, 2000). Effect size indices for situations where two or more grouping variables are used in various combinations of fixed and random factor effects are provided by Olejnik & Algina (2000) as well as Kline (2004a: Chapter 7).

2.4 **Effect size indices based on Group overlapping**

Tilton (1937) suggested that the comparison between two means should be, as far as is possible, enhanced by making use of a measure of overlapping such as the percentage of area under the distribution common to both distributions. Cohen (1977:23) defines various overlapping indices when two normally distributed populations with equal standard deviations are used and also provides the relationship of this quantity to the standardized difference in means (*d*). This index could, for example, be used to help interpret the value *d*: a small degree of overlapping suggests large values of *d* (for further reading see Huberty & Lowman, 2000, as well as Kline, 2004a: Figure 4.1).

In this way group overlapping can be related to forecast discriminant analysis. Huberty & Holmes (1983) apply it in the univariate two-sample case, where they relate the probability of correct classification (percentage of overlapping) to the effect size index *d*. A more sensible "better-than-chance-hit-rate" index (*I*) is suggested by Huberty (1994), while Huberty & Lowman (2000) made use of this idea in a multivariate context. Attention is paid to this topic in Chapter 8.

2.5 Multivariate Effect size indices

In the case where a dependent continuous variable's relationship with more than one independent variable has to be determined, (i.e., with multiple linear regression, abbreviated as MLR) the coefficient of multiple correlation (*R*) is the appropriate measure. In an attempt to relate MLR to ANOVA, Cohen (1977:410) introduced the f^2 index which describes the *explained variation* $\left(R^2\right)$ relative to the unexplained variation $\left(1-R^2\right)$. He goes further to define f^2 in terms of the *partial* $\boldsymbol{R}^2$ which in turn is a function of the *semi-partial* $\boldsymbol{R}^2$. The semi-partial $\boldsymbol{R}^2$ is defined as the increase in R^2 when a set of variables, B, is added to the existing set, A (see also Smithson, 2001).

In multivariate analysis of variance (MANOVA), Wilk's Λ is a well known statistic. The generalized $\eta^2 = 1-\Lambda$ is therefore a logical effect size index. Similarly, we could also use $\tau^2 = 1-\Lambda^{1/s}$, with *s = min (p, q),* where *p* is the number of variables and *q* is the hypothesized degrees of freedom (equal to *k-1* if *k* populations are compared.) (Rencher, 1995:192).

Other indices are ζ and ξ, which are based on the Hotelling-Lawley Trace statistic and Pillai's statistic (Rencher, 1995; Huberty, 1994; Olejnik & Algina, 2000). Modifications were also introduced to estimate the generalized η^2 (see Olejnik & Algina, 2000:273 and Steyn & Ellis, 2009).

In the case where only two populations are compared with one another, the *Mahalanobis distance,* D_M^2 , is used as an effect size index (Kline, 2004b: 3-4 and Steyn & Ellis, 2009)). This is a generalization of Cohen's *d* for more than one variable. In the case where one has more than two populations, D_M^2 can also be defined for *contrasts* and expressed in terms of Wilks' Λ for the contrasts. (See Kline, 2004b: 10).

The better-than-chance-hit-rate index by Huberty (1994) and Huberty & Lowman (2000) is, by its very nature, a multivariate index and these authors show how it can be used as an index for more than two groups.

2.6 Concluding remarks

Huberty (2002) provides the details and history of many other effect size indices, and goes on to summarize it graphically in a diagram of these indices along with dates (his Figure 1). The discussion up to this point has focused only on the most important indices. In the following chapters an attempt will be made to provide the full details of each of the effect sizes along with the methods used to calculate them and illustrative examples. When calculations are difficult to do by hand, an effort will be made to supply any available computer programs/packages/spread sheets which can be used to perform these calculations.