AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Heterogeneity coefficients for Mahalanobis' D as a multivariate effect size

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/1853046 since 2022-04-11T01:16:30Z

Published version:

DOI:10.1080/00273171.2016.1262237

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

20 April 2024

Heterogeneity Coefficients for Mahalanobis' *D* as a Multivariate Effect Size

Marco Del Giudice

University of New Mexico

***Multivariate Behavioral Research, 52,* 216-221 (2017).**

Marco Del Giudice, Department of Psychology, University of New Mexico.
Address correspondence to Marco Del Giudice, Department of Psychology, University of New Mexico. Logan Hall, 2001 Redondo Dr. NE, Albuquerque, NM 87131, USA; email: marcodg@unm.edu

Abstract

The Mahalanobis distance *D* is the multivariate generalization of Cohen's *d*, and can be used as a standardized effect size for multivariate differences between groups. An important issue in the interpretation of *D* is heterogeneity, that is, the extent to which contributions to the overall effect size are concentrated in a small subset of variables rather than evenly distributed across the whole set. Here I present two heterogeneity coefficients for *D* based on the Gini coefficient, a well-known index of inequality among values of a distribution. I discuss the properties and limitations of the two coefficients and illustrate their use by reanalyzing some published findings from studies of gender differences.

Keywords: effect size; group differences; heterogeneity; Mahalanobis distance; multivariate.

Introduction

In the behavioral sciences, the effect size of choice for differences between groups—for example males versus females, patients versus controls, or participants in experimental versus control conditions—is the standardized mean difference or Cohen's *d* (Cohen, 1988). While *d* is adequate for univariate differences, it cannot be used to properly quantify overall group differences in multidimensional domains such as personality, aggression, mate preferences, and so on. When investigating group differences in one of those domains, the standard practice of considering univariate *d*'s one at a time can be severely misleading, for two reasons. First, relatively small differences in many individual variables may easily add up to a large overall difference. Second, univariate effect sizes inevitably fail to take into account the pattern of correlation between the variables that make up the domain of interest. The need for standard, easily interpretable multivariate indices has become more acute with the ongoing shift from significance testing to effect size estimation that defines the so-called "new statistics" in psychology (Cumming, 2014).

The Mahalanobis Distance

In a series of papers, I have proposed the Mahalanobis distance *D* as the natural effect size for multivariate group differences (Del Giudice, 2009, 2011, 2013). Mahalanobis' *D* is the multivariate generalization of Cohen's *d*, and has the same substantive interpretation in terms of distribution overlap (see Bradley, 2006; Huberty, 2005; Reiser, 2001). Specifically, *D* is the (unsigned) standardized difference between the two groups along the discriminant axis; for example, $D = 1.00$ means that the group centroids are one standard deviation apart along the discriminant axis. If the variables are all orthogonal, *D* reduces to the Euclidean distance; if there is only one variable in the set, *D* reduces to Cohen's *d.* As with Cohen's *d*, Mahalanobis' *D* can be translated into approximate measures of statistical overlap by assuming multivariate normality (see Del Giudice, 2009). Confidence intervals on *D* are easy to obtain with bootstrapping (see Kelley, 2005); exact analytical methods are also available, although they are not always applicable (Reiser, 2001; Zou, 2007). While sample estimates of *D* are biased upward, the bias can be minimized by increasing sample size; initial simulations suggest that a ratio of about 100 cases per variable should be sufficient in most standard applications (Del Giudice, 2013).

The Mahalanobis distance was introduced eighty years ago (Mahalanobis, 1936) and is a standard tool in multivariate analysis. Surprisingly, though, it has not been used as an effect size in psychological research until very recently. Thanks to a widely disseminated study (Del Giudice, Booth, & Irwing, 2012), more researchers have started to employ *D* as a multivariate effect size to supplement and extend the standard univariate approach, particularly in relation to gender differences (e.g., Conroy-Beam, Buss, Pham, & Shackelford, 2015; Del Giudice, Lippa, Puts, Bailey, Bailey, & Schmitt, 2015; Morris, 2016; Vianello, Schnabel, Sriram, & Nosek, 2013).

To illustrate the usefulness of multivariate effect sizes one may consider the case of gender differences in personality. Reviews and meta-analyses on this topic have traditionally considered one personality trait at a time, and have employed the average of the univariate effects as an overall index of size. This has led to the widely shared idea that gender differences in personality are small and inconsistent, with a large overlap between the male and female distributions (see Hyde, 2005, 2014). However, the standard approach may have missed the forest for the trees. Individual personality profiles reflect the combination of multiple dimensions, and the same trait value may manifest in different ways depending on the level of other traits (think of a neurotic and agreeable person vs. a neurotic but disagreeable one; see Larsen & Buss, 2013). What defines the average profiles of men and women may not be a specific trait, but a particular combination of narrow tendencies across multiple traits. This is precisely what happens in the domain of facial morphology: while gender differences in individual anatomical features (e.g., nose length, eyebrow thickness) are generally small to moderate, their combination results in a large overall difference, with an overlap of less than 10% between the multivariate distributions in males and females (and $D \approx 3$). This separation enables observers to identify the gender of faces with very high accuracy (see Del Giudice, 2013). Consistent with this view, my colleagues and I found a multivariate effect size of $D = 2.71$ for gender differences in personality (Del Giudice et al., 2012; more details below). This corresponds to an overlap of only 10% between the male and female distributions (assuming multivariate normality). In other words, the overall personality profiles of males and females are in fact quite distinct, despite the considerable amount of variation within each gender. In this case, a multivariate perspective revealed a previously undescribed empirical pattern and challenged the existing consensus on an important psychological phenomenon.

The Issue of Heterogeneity

While D can provide valuable information on patterns of group differences, its interpretation is more complex than that of d, and raises additional issues that do not arise in the standard univariate framework. One of these issues is heterogeneity in the determination of D. Finding a large multivariate difference between two groups does not tell whether the overall effect size reflects (a) the joint contribution of many variables, or (b) the overwhelming contribution of one or a few variables. For example, the multivariate effect size in Del Giudice and colleagues (2012) was $D = 2.71$ in a set of 15 personality traits. However, removing just one trait (Sensitivity) from the set reduced the difference to $D = 1.71$, indicating a disproportionate contribution of gender differences in Sensitivity to the overall effect. As another case in point, consider the cross-cultural analysis of mate preferences by Conroy-Beam and colleagues (2015). These authors used D to quantify gender differences in preferences for traits such as good looks, health, and sociability in romantic partners. They found that the overall size of the differences between men and women becomes smaller in countries with higher levels of gender equality. Still, their analysis does not reveal whether the reduction in D observed in gender-egalitarian countries reflects smaller differences across the board or—as some data in the study suggest—a disproportionate influence of a few specific traits (e.g., ambition) in the more unequal societies. The first scenario implies a more or less constant amount of heterogeneity across countries; the second scenario implies a decrease in heterogeneity as one moves toward more gender-egalitarian countries. While the size of univariate differences can point researchers toward potentially influential variables, the net contribution of a given variable also depends on its correlations with the other variables in the set, and cannot be judged adequately by simply relying on univariate indices.

In this paper I tackle this problem by proposing two simple heterogeneity coefficients for *D*. Both coefficients are based on the Gini coefficient, a well-known index of inequality among values of a distribution. The heterogeneity coefficient *H* quantifies inequality in the contribution of individual variables to the overall effect size, and ranges from 0 (maximal homogeneity) to 1 (maximal heterogeneity). Specifically, a value of *H* = 1 indicates that the multivariate effect is fully determined by the positive contribution of just one variable. The complementary coefficient *EPV* (for *equivalent proportion of variables*) expresses *H* in a more concrete form, as the proportion of equally contributing variables that would produce the same amount of heterogeneity if the remaining variables in the set made no contribution. These indices of heterogeneity can aid researchers in the interpretation of *D*, and provide finer-grained information on the nature of multivariate group differences. After deriving formulas for *H* and *EPV*, I illustrate their use by reanalyzing some published findings from studies of gender differences in psychology and neuroscience.

Heterogeneity Coefficients for *D*

The *H* Coefficient

For a set of *n* variables measured in two groups *A* and *B*, the Mahalanobis distance *D* is given by

$$D = [(\mathbf{m}_A - \mathbf{m}_B)^{\mathrm{T}} \mathbf{S}^{-1} (\mathbf{m}_A - \mathbf{m}_B)]^{\frac{1}{2}} \quad (1)$$

where $\mathbf{m}_A$ and $\mathbf{m}_B$ are column vectors of the means of the *n* variables in groups *A* and *B*, and **S** is the common variance-covariance matrix. Equivalently,

$$D = (\mathbf{d}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{d})^{\frac{1}{2}} \quad (2)$$

where **d** is a column vector of standardized differences (Cohen's *d*) and **R** is the common correlation matrix. From Eq. 2, the squared Mahalanobis distance D^2 can be written as the sum

$$D^2 = \mathbf{d}^{\mathrm{T}} \mathbf{Z} \mathbf{d} = \sum_{i=1}^{n} \sum_{j=1}^{n} z_{ji} d_j d_i = \sum_{i=1}^{n} C_i, \quad (3)$$

where $\mathbf{Z} = \mathbf{R}^{-1}$ and

$$C_i = \sum_{j=1}^{n} z_{ji} d_j d_i. \quad (4)$$

When $d_i = 0$, $C_i = 0$. When $d_i \neq 0$ (that is, for nonzero univariate differences), C_i can be rewritten as:

$$C_i = \left(\sum_{j=1}^{n} z_{ji} \frac{d_j}{d_i}\right) d_i^2. \quad (5)$$

In other words, the squared Mahalanobis distance can be decomposed into a weighted sum of squared univariate differences. The weight $\left(\sum_{j=1}^{n} z_{ji} \frac{d_j}{d_i}\right)$ in Eq. 5 captures the effect of the correlations between the variable of interest and the other variables. If all variables are orthogonal, **Z** is an identity matrix and $C_i = d_i^2$; the Mahalanobis distance then reduces to the Euclidean distance so that

$$D^2 = \sum_{i=1}^{n} d_i^2. \tag{6}$$

Conveniently, C_i can be interpreted as the net contribution of each variable to the multivariate effect size. If the effect size d_i is zero, the variable makes no contribution. Otherwise, the variable's contribution corresponds to the squared effect size (d_i^2) weighted by a term that is a function of its correlations with the other variables and may be positive, zero, or negative.

A heterogeneity coefficient for *D* should be minimized when all the variables make the same contribution ($C_1 = C_2 = \cdots = C_n$), and maximized when one variable fully accounts for the size of the multivariate effect. The Gini coefficient *G* offers a straightforward way to quantify inequality in the contributions of the *n* variables. When the values used to compute *G* are non-negative, the coefficient has a lower bound of 0 (maximum equality) and approaches 1 (maximum inequality) as the number of values becomes very large; for *n* values, the upper bound on *G* is $(n-1)/n$. However, the coefficient in the standard form can become larger than 1 if the distribution contains negative values, which is problematic since C_i values can be negative (see Raffinetti, Siletti, & Vernizzi, 2015). While *G* can be normalized to accommodate negative values (Berrebi & Silber, 1985; Chen, Tsaur, & Rhai, 1982; Raffinetti et al., 2015), the normalized coefficient has an undesirable property: its upper bound is no longer $(n-1)/n$ but depends on the exact distribution of positive and negative values. As a consequence, normalized Gini coefficients are hard to interpret and compare across studies.

For the purpose of deriving a heterogeneity coefficient for *D*, an effective solution is to focus on the *positive* contributions to the effect size, by treating the sum of the negative contributions (if any) as the "baseline" to which the remaining variables add to obtain the total effect size. The new quantity of interest is not D^2 but rather the sum of the positive contributions to D^2, which is larger than D^2 if some C_i values are negative. This is obtained by setting negative C_i values to zero. This way, variables with negative C_i values still count toward increasing heterogeneity, but the problems with the normalized Gini coefficient can be avoided. The recoded C_i values are indicated as C_i^*, so that

$$C_i^* = max(0, C_i). \tag{7}$$

It thus becomes possible to apply the standard formula for the Gini coefficient:

$$G = \frac{(2/n^2)\sum_{i=1}^{n} i x_i - [(n+1)/n^2]\sum_{i=1}^{n} x_i}{\bar{x}} \tag{8}$$

where $x_1 \dots x_n$ are *n* ordered values ($x_1 < x_2 \dots < x_n$) and $\bar{x}$ is their average (see Raffinetti et al., 2015). A correction can be applied by multiplying the expression in Eq. 8 by $n/(n-1)$, thus ensuring that *G* always has an upper bound of 1 even when *n* is small (Deltas, 2003). Replacing x with C^* in Eq. 8 and applying the correction yields the heterogeneity coefficient *H*:

$$H = \frac{(2/n)\sum_{i=1}^{n} iC_i^* - [(n+1)/n]\sum_{i=1}^{n} C_i^*}{(n-1)\bar{C}^*} \tag{9}$$

where n is the number of variables, $C_1^* \ldots C_n^*$ are the ordered values of C_i^* (with $C_1^* < C_2^* \ldots < C_n^*$) and $\bar{C}^*$ is their average. The heterogeneity coefficient in Eq. 9 ranges from $H = 0$ when all the variables in the set contribute equally, to $H = 1$ when the size of the multivariate effect entirely depends on the (positive) contribution of one variable.

The *EPV* Coefficient

The *EPV* coefficient is the proportion of equally contributing variables that would produce the same amount of heterogeneity, if the remaining variables in the set made no contribution. *EPV* equals one minus the uncorrected Gini coefficient (Eq. 8), which corresponds to:

$$EPV = 1 - \frac{n-1}{n}H. \tag{10}$$

For example, $H = .67$ with $n = 10$ variables corresponds to $EPV = .40$; this means that the same level of heterogeneity would obtain in a hypothetical scenario where 40% of the variables contributed equally to *D* and the remaining 60% made no contribution.

The formula for *EPV* follows directly from the definition of the Gini coefficient as the ratio of (a) the area between the line of equality and the Lorenz curve and (b) the total area below the equality line (see e.g., Chen et al., 1982). If a proportion p of the variables contribute equally while the remaining ones make no contribution, the area below the Lorenz curve is a proportion p of the total area below the equality line, and the uncorrected Gini coefficient becomes $1 - p$ by definition. The *EPV* coefficient approaches $1 - H$ as n becomes larger, but the two values may diverge considerably when the number of variables is small.

While *EPV* may provide a more intuitive summary of heterogeneity than *H*, its main limitation is that it can never be smaller than $1/n$ (when $H = 1$). For this reason, one should be cautious when comparing *EPV* values calculated on markedly different numbers of variables. When there are at least 5 variables in the set, $EPV \leq .20$ may be used as a reasonable criterion to flag high levels of heterogeneity (corresponding to a scenario in which 20% or less of the variables explain 100% of the effect size).

Empirical Illustrations

In the personality study cited above, my colleagues and I found a multivariate effect size of $D = 2.71$ on a set of 15 personality traits (mean differences and correlations where estimated via multigroup latent variable modeling). However, one particular variable (Sensitivity) showed a large univariate difference ($d = 2.29$), suggesting a disproportionate contribution to the overall effect (see Del Giudice et al., 2012). Indeed, the heterogeneity coefficients for this effect size are $H = .95$ and $EPV = .11$, indicating a very high level of heterogeneity. Removing the Sensitivity variable from the analysis reduced the effect size to $D = 1.71$; the corresponding heterogeneity coefficients are $H = .80$ and $EPV = .25$. The new value of *EPV* is noticeably larger than before, and lies above the proposed .20 cutoff for high heterogeneity. It should be stressed that these results do not imply that the original effect size $D = 2.71$ is somehow "invalid" or should be discarded in favor of $D = 1.71$. This kind of decision depends on the specific theoretical and practical goals of a given study; depending on context, a high level of heterogeneity may well be expected and/or desirable. What these results clearly show is that patterns of sex differences are not evenly distributed across personality traits; instead, they tend to be concentrated in a relatively small subset of traits. Also, the .20 cutoff is merely a rule of thumb to aid interpretation, and should not be reified or employed blindly to categorize *EPV* values into "small" and "large" regardless of context.

In another large-scale study, Morris (2016) analyzed the size of gender differences in occupational preferences, measured on six standard dimensions (realistic, investigative, artistic, social, enterprising, and conventional). In the total sample, the multivariate effect was $D = 1.70$ (corrected for attenuation). The heterogeneity coefficients for this effect are $H = .87$ and $EPV = .27$; these values indicate that gender differences in occupational preferences are concentrated in a relatively small subset of dimensions (M. Morris, personal communication: October 10, 2016).

Aggression is another important domain in which males and females show systematic differences. In Del Giudice (2009) I reanalyzed Archer's (2009) meta-analytic findings on physical, verbal, and indirect aggression. Correcting for attenuation, the multivariate effect size computed on the three domains of aggression could be estimated at about $D = 0.89-1.01$, depending on assumptions about correlation patterns. The corresponding heterogeneity coefficients are $H = .33-.63$ and $EPV = .58-.78$. The amount of heterogeneity observed in this case corresponds to a scenario in which about 70% of the variables contribute equally to *D*. These values indicate lower heterogeneity than in the case of personality or occupational preferences. However, one should keep in mind that, with $n = 3$, *EPV* can never be smaller than .33; thus, it would be misleading to directly compare the *EPV* of gender differences in 15 personality traits to that of gender differences in 3 domains of aggression. (Also, note that the proposed .20 cutoff is meaningless with fewer than 5 variables.) The corresponding *H* coefficients are more comparable, as they have the same range regardless of the number of variables considered (i.e., $H = 1.00$ always means that one variable explains the totality of the effect size).

Finally, my colleagues and I (Del Giudice et al., 2015) reanalyzed the data on gender differences in brain anatomy presented by Joel and collaborators (2015). The six datasets we reanalyzed contained information of three types of anatomical features—volume, cortical thickness, and fractional anisotropy (a measure of white matter integrity). From each dataset, Joel and collaborators (2015) had selected 7−12 variables, each measuring a feature (e.g., volume) of a specific brain region. Across datasets, multivariate effect sizes computed using the same variables ranged from $D = 0.69$ to $D = 1.47$. The corresponding heterogeneity coefficients range from $H = .44$ and $EPV = .58$ (in the cortical thickness dataset labeled NKI, SBA) to $H = .70$ and $EPV = .36$ (in the fractional anisotropy dataset; for details see Joel et al., 2015). The unweighted averages across the six datasets are .56 for *H* and .50 for *EPV*. While none of the datasets shows high levels of heterogeneity according to the .20 cutoff, the fairly broad range of *H* values suggests the possibility that some aspects of gender differences in brain anatomy may be more strongly localized (i.e., concentrated in a few specific regions) than others. These findings illustrate how heterogeneity coefficients may have substantive implications for empirical research.

Conclusion

As multivariate effect sizes such as *D* become more widely used in research on group differences, they will necessitate new statistical tools to aid in their interpretation, diagnose potential problems, and so on. Here I presented two simple coefficients that quantify heterogeneity in the contribution of individual variables to the overall effect size. The *H* coefficient always ranges between 0 and 1, regardless of the number of variables considered. The *EPV* coefficient expresses heterogeneity in more intuitive terms, but its range inevitably depends on the number of variables in the set, which may complicate interpretation and limit the ability to make comparisons between studies.

While all the empirical examples I discussed in this paper concern gender differences, the potential applications of *D* are much broader, and by no means limited to correlational studies. As a hypothetical experimental scenario, consider a study investigating the effects of various treatments on multiple dimensions of psychotic symptoms (for example positive symptoms such as hallucinations, negative symptoms such as anhedonia, and disorganization symptoms such as psychomotor agitation). Different treatments may produce different patterns of change—some may lead to larger symptom reductions limited to a specific dimension, while others may result in smaller but more homogeneous improvements. Moreover, a treatment associated with relatively small improvements in each dimension may still produce the largest overall benefit when all the variables are considered simultaneously. These multivariate patterns would be easy to summarize and compare using *D* in combination with heterogeneity indices such as *H*.

To facilitate those who wish to use these coefficients in their research, an R script can be downloaded at http://marcodg.net/publications or obtained directly from the author. The functions included in the script compute Mahalanobis' *D* from raw data or summary statistics and return confidence intervals, overlap coefficients, heterogeneity coefficients, and other useful indices.

- See Addendum Below -

References

Archer, J. (2009). Does sexual selection explain human sex differences in aggression? *Behavioral and Brain Sciences, 32*, 249–266. doi:10.1017/S0140525X09990951

Bradley, E. L. (2006). Overlapping coefficient. In S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic (Eds.), *Encyclopedia of Statistical Sciences* (2nd ed.) (pp. 546-547). Hoboken, NJ: Wiley.

Berrebi, Z. M., & Silber, J. (1985) The Gini coefficient and negative income: A comment. *Oxford Economic Papers, 37,* 525– 526.

Chen, C. N., Tsaur, T. W., & Rhai, T. S. (1982) The Gini coefficient and negative income. *Oxford Economic Papers, 34,* 473–478.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Conroy-Beam, D., Buss, D. M., Pham, M. N., & Shackelford, T. K. (2015). How sexually dimorphic are human mate preferences? *Personality and Social Psychology Bulletin, 41,* 1082-1093. doi:10.1177/0146167215590987

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25,* 7-29. doi:10.1177/0956797613504966

Del Giudice, M. (2009). On the real magnitude of psychological sex differences. *Evolutionary Psychology, 7,* 264-279. doi:10.1177/147470490900700209

Del Giudice, M. (2011). Sex differences in romantic attachment: A meta-analysis. *Personality and Social Psychology Bulletin, 37,* 193-214. doi:10.1177/0146167210392789

Del Giudice, M. (2013). Multivariate misgivings: Is D a valid measure of group and sex differences? *Evolutionary Psychology, 11,* 1067-1076. doi:10.1177/147470491301100511

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS ONE, 7,* e29265. doi:10.1371/journal.pone.0029265

Del Giudice, M., Lippa, R. A., Puts, D. A., Bailey, D. H., Bailey, J. M., & Schmitt, D. P. (2015). *Mosaic brains? A methodological critique of Joel et al. (2015).* doi:10.13140/RG.2.1.1038.8566

Deltas, G. (2003), The small-sample bias of the Gini coefficient: Results and implications for empirical research. *Review of Economics and Statistics, 85,* 226-234. doi:10.1162/rest.2003.85.1.226

Huberty, C. J. (2005). Mahalanobis distance. In B. S. Everitt and D. C. Howell (eds.), *Encyclopedia of statistics in behavioral science.* Chichester: Wiley.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581-592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373-398. doi:10.1146/annurev-psych-010213-115057

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., ... & Liem, F. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences USA, 112,* 15468-15473. doi:10.1073/pnas.1509654112

Kelley, K. (2005). The effects of nonnormal distributions on confidence intervals around the standardized mean difference: Bootstrap and parametric confidence intervals. *Educational and Psychological Measurement, 65,* 51-69. doi:10.1177/0013164404264850

Larsen, R. J., & Buss, D. M. (2013). *Personality psychology: Domains of knowledge about human nature* (5th ed.). New York: McGraw-Hill.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science, Calcutta, 2,* 49-55.

Morris, M. L. (2016). Vocational interests in the United States: Sex, age, ethnicity, and year effects. *Journal of Counseling Psychology*. doi:10.1037/cou0000164

Raffinetti, E., Siletti, E., & Vernizzi, A. (2015). On the Gini coefficient normalization when attributes with negative values are considered. *Statistical Methods and Applications, 24,* 507-521. doi:10.1007/s10260-014-0293-4

Reiser, B. (2001). Confidence intervals for the Mahalanobis distance. *Communications in Statistics: Simulation and Computation, 30,* 37–45. doi:10.1081/SAC-100001856

Vianello, M., Schnabel, K., Sriram, N., & Nosek, B. (2013). Gender differences in implicit and explicit personality traits. *Personality and Individual Differences, 55,* 994-999. doi:10.1016/j.paid.2013.08.008

Zou, G. Y. (2007). Exact confidence interval for Cohen's effect size is readily available. *Statistics in Medicine, 26,* 3054 doi:10.1002/sim.2830

Addendum (Multivariate Behavioral Research, 53, 571-573 [2018]).

Multivariate effect sizes such as Mahalanobis' *D* raise the issue of heterogeneity in the contributions of individual variables to the overall effect. In Del Giudice (2017) I proposed a strategy to quantify heterogeneity: first, partition D^2 into a set of non-negative values that reflect the contributions of individual variables; second, apply the small-sample Gini formula to those values to obtain a heterogeneity coefficient ranging between 0 and 1. The critical step in this strategy is finding a suitable partition of D^2. In Del Giudice (2017) I used a simple approach to partition the multivariate D^2 into a weighted sum of the squared univariate effects (d_i^2). The resulting C_i values have two desirable properties. First, $C_i = 0$ if removing variable X_i from the set leaves *D* unchanged. For two variables X_1, X_2 with correlation *r* and effect sizes d_1, d_2 (the case I will use for illustration here):

$$C_1 = \frac{1}{1-r^2}\left(1 - r\frac{d_2}{d_1}\right) d_1^2 \qquad (1)$$

And

$$C_2 = \frac{1}{1-r^2}\left(1 - r\frac{d_1}{d_2}\right) d_2^2. \qquad (2)$$

It follows that $C_1 = 0$ when $d_1 = rd_2$, which is appropriate since in this case $D^2 = (C_1 + C_2) = d_2^2$. Conversely, $C_2 = 0$ when $d_2 = rd_1$, and $D^2 = d_1^2$. Second, when two highly correlated variables X_i and X_j are both in the set, their joint contribution is split between C_i and C_j, and is not partialed out as it would happen with methods for quantifying the contribution of individual variables that proceed by removing one variable at a time (e.g., Rencher, 1993).

In the original paper, I suggested that C_i values can be interpreted as "net contributions" to D^2. This is incorrect: C_i values can be negative, even though D^2 never decreases when more variables are added to the set. The fact that C_i values can be negative also makes them unsuitable for calculating the standard Gini coefficient. In Del Giudice (2017), I proposed an ad-hoc solution to this problem, namely, setting negative C_i values to zero and using the resulting C_i^* values to calculate two Gini-based heterogeneity coefficients, *H* and *EPV*. However, this approach is not ideal, because a negative C_i value still reflects a positive contribution of X_i to *D* (that is, *D* decreases if X_i is removed from the set). To illustrate, in the two-variable case with $r > 0$, negative values $C_i < 0$ occur whenever $0 < d_i < rd_j$; it is easy to show that under the same conditions $D^2 = \left(C_i + C_j\right) > d_j^2$, which implies that X_i makes a positive contribution to *D*. Negative C_i values are more likely to occur when d_i is small relative to d_j and the two variables are strongly correlated. In sum, setting negative C_i values to zero is an overly conservative approach—it tends to underestimate the contribution of some variables to *D*, and hence overestimate the amount of heterogeneity. The resulting distortion tends to become larger as collinearity among the variables increases.

A better solution to the problem of negative C_i values is to use the ordered absolute values $|C_1| \dots |C_n|$ (where $|C_1| < |C_2| \dots < |C_n|$, and $\overline{|C|}$is their average) to calculate heterogeneity. To avoid confusion, the resulting coefficients can be labeled as H_2 and EPV_2:

$$H_2 = \frac{(2/n)\sum_{i=1}^{n} i|C_i| - [(n+1)/n]\sum_{i=1}^{n}|C_i|}{(n-1)\overline{|C|}} \tag{3}$$

and

$$EPV_2 = 1 - \frac{n-1}{n} H_2. \tag{4}$$

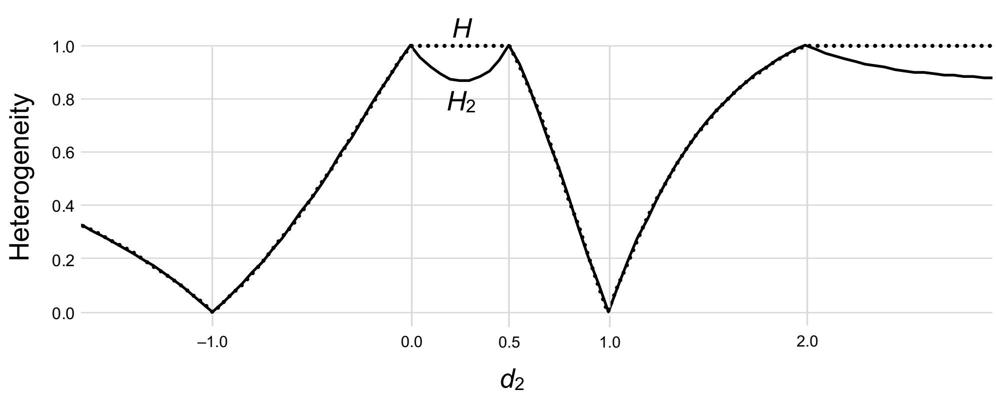

Figure 1. Behavior of heterogeneity coefficients *H* and H_2 in the case of two positively correlated variables. (In the example shown, $d_1 = 1.0$ and $r = .5$; the qualitative pattern does not depend on the choice of values.) The dotted line shows the original *H* coefficient. The solid line shows the revised H_2 coefficient discussed here. For both coefficients, heterogeneity is minimal when $d_2 = \pm d_1$ (in this example, $d_2 = \pm 1$) and maximal when $d_2 = 0$, when $d_2 = rd_1$ (in this example, $d_2 = 0.5$), and when $d_2 = d_1/r$ (in this example, $d_2 = 2$).

Figure 1 illustrates the behavior of H and H_2 in the two-variable case (with $r, d_1 > 0$). For both $0 < d_2 < rd_1$ and $d_2 > d_1/r$ (equivalent to $0 < d_1 < rd_2$), coefficient H remains equal to 1 (maximum heterogeneity, wrongly indicating that only one variable contributes to *D*), whereas H_2 correctly decreases to reflect the nonzero contributions of d_2 and d_1, respectively.

To illustrate the difference between H and H_2 with some real-world examples, consider the empirical datasets analyzed in Del Giudice (2017). For the aggression dataset from Del Giudice (2009), using H or H_2 makes no difference because there are no negative C_i values. For the personality dataset from Del Giudice, Booth, & Irwing (2012), $H = .95$ ($EPV = .11$) while $H_2 = .90$ ($EPV_2 = .16$). After removing the "sensitivity" factor from the analysis, $H = .80$ ($EPV = .25$) while $H_2 = .76$ ($EPV_2 = .30$). For the brain anatomy data discussed in Del Giudice et al. (2016), H ranges from .44 to .70 (EPV from .36 to .58); the corresponding values of H_2 range from .32 to .58 (EPV_2 from .47 to .71). In these examples, H_2 does not dramatically change the picture, but it does suggest a somewhat more homogeneous contribution than indicated by H.

While coefficient H_2 improves on the original H, both have limitations owing to their reliance on the C partition. In particular, $C_i = 0$ whenever $d_i = 0$; however, a variable X_i may contribute to increase D even if $d_i = 0$, provided that it has nonzero correlations with the other variables. In the two-variable case, it is easy to show that if $d_i = 0$ and $r \neq 0$, then $D^2 > d_j^2$. Future research may show a way to partition D^2 so as to avoid this problem while maintaining the desirable properties of C. The currently available alternatives are not well suited for the task—for example, the partitioning method recently proposed by Garthwaite & Koch (2016) avoids the problem of negative values, but fails to identify cases in which one of the variables makes no contribution to D (in the two-variable case, when $d_2 = rd_1$ or $d_2 = d_1/r$). At present, H_2 offers a practical means to quantify heterogeneity and should prove useful in a variety of applications.

Acknowledgments

I am much indebted to Benjamin Reiser for pointing out some limitations of my original paper, suggesting additional literature, and commenting on a draft of this addendum.

References

Del Giudice, M. (2009). On the real magnitude of psychological sex differences. *Evolutionary Psychology, 7,* 264-279. doi:10.1177/147470490900700209

Del Giudice, M. (2017). Heterogeneity coefficients for Mahalanobis' *D* as a multivariate effect size. *Multivariate Behavioral Research, 52,* 216-221. doi:10.1080/00273171.2016.1262237

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLoS ONE, 7,* e29265. doi:10.1371/journal.pone.0029265

Del Giudice, M., Lippa, R. A., Puts, D. A., Bailey, D. H., Bailey, J. M., & Schmitt, D. P. (2016). Joel et al.'s method systematically fails to detect large, consistent sex differences.

Proceedings of the National Academy of Sciences USA, 113, E1965. doi:10.1073/pnas.1525534113

Garthwaite, P. H., & Koch, I. (2016). Evaluating the contributions of individual variables to a quadratic form. *Australian & New Zealand Journal of Statistics, 58,* 99-119. doi:10.1111/anzs.12144

Rencher, A. C. (1993). The contribution of individual variables to Hotelling's T^2, Wilks' Λ, and R^2. *Biometrics, 49,* 479-489. doi:10.2307/2532560