maha_v4.R

Two simple R functions that compute Mahalanobis' *D*, confidence intervals, overlap coefficients, and heterogeneity coefficients, either from raw data with **maha()** or from Cohen's *d* values and a common correlation matrix with **maha.summary()**.For general information see Del Giudice (2009, 2013).

By Marco Del Giudice (2016). Version 4.

Note on confidence intervals: exact confidence intervals are computed with Reiser's (2001) method. Especially for small D values, the equations may not be solvable; in those cases, one or both CI bounds are set to NA. For more information see Reiser (2001). Bootstrapped CIs are bias-corrected and accelerated; for details see Kelley (2005).

Note on heterogeneity coefficients: the heterogeneity coefficient H quantifies heterogeneity in the variables' contribution to the multivariate effect size (0 = max homogeneity; 1 = max heterogeneity). The EPV coefficient (equivalent proportion of variables) is the proportion of contributing variables that would result in the same heterogeneity, in a hypothetical scenario where a certain proportion of variables contribute equally to D while the remaining ones make no contribution. See Del Giudice (in press - 2016).

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. (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. (in press - 2016). Heterogeneity coefficients for Mahalanobis' D as a multivariate effect size. Multivariate Behavioral Research.

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

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

maha(data_A, data_B, alpha=NULL, conf.level=.95, boot.n=10000)

Returns the Mahalanobis distance D, confidence intervals, heterogeneity coefficients, and coefficients of overlap, computed from raw data. Can compute disattenuated estimates if desired. Note: The correlation matrices of the two groups are pooled by taking weighted averages before computing D. Tucker's Congruence Coefficient provides an index of similarity between the two correlation matrices (0 = min similarity; 1 = max similarity).

Arguments
data_A, data_B: raw matrices/data frames for the two groups.
alpha: vector of reliability coefficients (optional: only required for disattenuation)
conf.level: CI width (optional)
boot.n: number of bootstrap samples (optional; deafult is 10,000)

Value
returns a list object containing some or all of the following:
D: mahalanobis *D*
CI_lower_exact: exact CI lower bound (NA if not solvable)
CI_upper_exact: exact CI upper bound (NA if not solvable)
CI_lower_boot: bootstrapped CI lower bound
CI_upper_boot: bootstrapped CI upper bound
OVL: coefficient of overlap (single distribution)
OVL2: Cohen's coefficient of overlap 1-U_1 (joint distribution)
CC_cor: Tucker's Congruence Coefficient (similarity between correlation matrices, 0-1)
H: heterogeneity coefficient *H* (0 = max homogeneity; 1 = max heterogeneity)
EPV: *EPV* coefficient (equivalent proportion of variables, 0-1)
d_values: vector of Cohen's *d* values
Dc: disattenuated *D*
OVLc: disattenuated coefficient of overlap (single distribution)
OVL2c: disattenuated Cohen's coefficient of overlap (joint distribution)
Hc: heterogeneity coefficient for disattenuated *D*
EPVc: *EPV* coefficient for disattenuated *D*
dc_values: vector of disattenuated Cohen's *d* values

maha.summary(d_values, cor_matrix, alpha=NULL, nA=NULL, nB=NULL, conf.level=.95)

Returns the Mahalanobis distance D, confidence intervals, heterogeneity coefficients, and coefficients of overlap, computed from summary statistics. Can compute disattenuated estimates if desired.

Arguments
d_values: row vector of Cohen's *d* values
cor_matrix: correlation matrix
alpha: vector of reliability coefficients (optional: only required for disattenuation)
nA, nB: sample size of the two groups (optional: only required for exact CI)
conf.level: CI width (optional)

Value
returns a list object containing some or all of the following:
D: mahalanobis *D*
CI_lower_exact: exact CI lower bound (NA if not solvable)
CI_upper_exact: exact CI upper bound (NA if not solvable)
OVL: coefficient of overlap (single distribution)
OVL2: Cohen's coefficient of overlap 1-U_1 (joint distribution)
H: heterogeneity coefficient *H* (0 = max homogeneity; 1 = max heterogeneity)
EPV: *EPV* coefficient (equivalent proportion of variables, 0-1)
d_values: vector of Cohen's *d* values
Dc: disattenuated *D*
OVLc = disattenuated coefficient of overlap (single distribution)
OVL2c: disattenuated Cohen's coefficient of overlap (joint distribution)
Hc: heterogeneity coefficient for disattenuated *D*
EPVc: *EPV* coefficient for disattenuated *D*
dc_values: vector of disattenuated Cohen's *d* values