maha_v8c.R

Marco Del Giudice (2018). Version 8c. Contact: marcodg@unm.edu

Two simple R functions to compute multivariate standardized differences between two groups (Mahalanobis' *D*), either from raw data with function `maha` or from Cohen's *d* values and correlations with function `maha.summary`. The functions return uncorrected and bias-corrected values of *D*, exact confidence intervals, bootstrapped confidence intervals (only from raw data), heterogeneity coefficients, and a number of additional diagnostics and effect sizes. Disattenuated estimates can also be computed if desired. For more information see Del Giudice (2009, 2013, 2019).

Confidence intervals: exact confidence intervals for *D* 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). The bootstrapped CIs are bias-corrected and accelerated; for details see Kelley (2005).

Additional effect sizes: additional effect sizes are calculated from *D* under the assumptions of multivariate normality and equality of covariance matrices. These are: the overlapping coefficient *OVL*; Cohen's alternative coefficient of overlap OVL_2 calculated on the joint distribution (equivalent to $1–U_1$); the common language effect size *CL* (McGraw & Wong, 1992); and the probability of correct classification (*PCC*) based on linear discriminant analysis (assuming equal group sizes). For details and discussion see Del Giudice (2019).

Heterogeneity coefficients: the heterogeneity coefficient H_2 quantifies heterogeneity in the variables' contribution to the multivariate effect size (0 = max homogeneity; 1 = max heterogeneity). Coefficient EPV_2 represents heterogeneity as the proportion of contributing variables that would result in the same heterogeneity (equivalent proportion of variables), in a hypothetical scenario where a certain proportion of variables contribute equally to *D* while the remaining ones make no contribution (range: 0-1). See Del Giudice (2017, 2018).

Similarity between correlation matrices: before computing *D* and other effect sizes, the correlation matrices of the two groups are pooled by taking weighted averages, unless a common correlation matrix is directly provided to function `maha.summary`. If two matrices are provided to function `maha.summary` without specifying the size of the groups, the unweighted average is taken. Tucker's Congruence Coefficient (*CC*) can be used as an index of similarity between the two sample correlation matrices (–1 = max dissimilarity; 1 = max similarity). Values above .85 indicate fair similarity; values above .95 indicate high similarity (Lorenzo-Seva & ten Berge, 2006). The significance of the difference between correlation matrices can be tested with Steiger's test of equality (Steiger, 1980); however, significance tests can be overly sensitive to small differences when sample size is large. See Del Giudice (2019) for more details.

Bias correction: the bias-corrected distance D_u removes the upward bias in *D*, which is especially large when sample size is small relative to the number of variables (see Hess et al., 2007; Del Giudice, 2019). The bias-corrected version of Cohen's *d* is d_u (also known as Hedges' *g*); the approximate formula is used here (see Kelley, 2005).

Disattenuation: disattenuated estimates of *D* and other effect sizes can be obtained by supplying a vector of reliability coefficients (e.g., Cronbach's α). For details see Del Giudice (2019).

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. (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. (2018). Addendum to: Heterogeneity coefficients for Mahalanobis' *D* as a multivariate effect size. *Multivariate Behavioral Research, 53,* 571-573. doi:10.1080/00273171.2018.1462138

Del Giudice, M. (2019). Measuring sex differences and similarities. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research*. New York, NY: Springer.

Hess, M. R., Hogarty, K. Y., Ferron, J. M., & Kromrey, J. D. (2007). Interval estimates of multivariate effect sizes: Coverage and interval width estimates under variance heterogeneity and nonnormality. *Educational and Psychological Measurement, 67,* 21-40. doi:10.1177/0013164406288159

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

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2,* 57-64. doi:10.1027/1614-2241.2.2.57

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111,* 361-365.

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

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research, 15,* 335-352. doi:10.1207/s15327906mbr1503_7

Function `maha`

Computes the Mahalanobis distance D, confidence intervals, and various relevant indices from raw data. Can compute disattenuated estimates if desired.

```
maha(dataA, dataB, rel_values=NULL, conf.level=.95, boot.n=NULL, round.digits=3)
```

Arguments

`dataA, dataB`	raw data matrices/data frames for the two groups
`rel_values`	vector of reliability coefficients (optional: required for disattenuation)
`conf.level`	CI width (default is 95%)
`boot.n`	number of bootstrap samples (optional; recommended: 5,000-10,000)
`round.digits`	rounding digits for the output (default is 3)

Value

returns a list object containing some or all of the following:

`D`	Mahalanobis' *D*
`CI_exact_D`	exact CI for *D* (lower bound, upper bound); NA if not solvable
`CI_boot_D`	bootstrapped CI for *D* (lower bound, upper bound)
`OVL`	overlapping coefficient *OVL* (single distribution)
`OVL2`	Cohen's coefficient of overlap OVL_2 (joint distribution; equals 1-U_1)
`CL`	common language effect size *CL*
`PCC`	probability of correct classification (*PCC*; linear discriminant analysis)
`H2`	heterogeneity coefficient H_2
`EPV2`	EPV_2 coefficient (equivalent proportion of variables)
`CC_cor`	Tucker's Congruence Coefficient *CC* (similarity between correlation matrices)
`steiger.p`	*p* value for Steiger's test of equality of correlation matrices (null hypothesis: equality)
`d_values`	vector of Cohen's *d* values

`Du`	bias-corrected D_u
`CI_boot_Du`	bootstrapped CI for D_u (lower bound, upper bound)
`OVLu`	overlapping coefficient (single distribution) based on D_u
`OVL2u`	Cohen's coefficient of overlap (joint distribution) based on D_u
`CLu`	common language effect size based on D_u
`PCCu`	probability of correct classification based on D_u
`du_values`	vector of bias-corrected d_u values (equivalent to Hedges' g)
`Dc`	disattenuated D_c
`OVLc`	disattenuated overlapping coefficient (single distribution)
`OVL2c`	disattenuated Cohen's coefficient of overlap (joint distribution)
`CLc`	disattenuated common language effect size
`PCCc`	disattenuated probability of correct classification
`H2c`	disattenuated heterogeneity coefficient H_2
`EPV2c`	disattenuated EPV_2 coefficient
`dc_values`	vector of disattenuated Cohen's d_c values
`Dcu`	disattenuated, bias-corrected D_{cu}
`OVLcu`	disattenuated overlapping coefficient (single distribution) based on D_{cu}
`OVL2cu`	disattenuated Cohen's coefficient of overlap (joint distribution) based on D_{cu}
`CLcu`	disattenuated common language effect size based on D_{cu}
`PCCcu`	disattenuated probability of correct classification based on D_{cu}
`dcu_values`	vector of disattenuated, bias-corrected d_{cu} values (disattenuated Hedges' g)

Function `maha.summary`

Computes the Mahalanobis distance D, confidence intervals, and various relevant indices from summary statistics. Can compute disattenuated estimates if desired.

```
maha.summary(d_values, corA, corB=NULL, nA=NULL, nB=NULL, rel_values=NULL,
       conf.level=.95, round.digits=3)
```

Arguments

`d_values`	(row) vector of Cohen's *d* values
`corA`	either the common correlation matrix, or the correlation matrix for group A
`corB`	correlation matrix for group B (optional)
`nA, nB`	sample size of the two groups (optional: required for bias correction, exact CIs, and Steiger's test)
`rel_values`	vector of reliability coefficients (optional: required for disattenuation)
`conf.level`	CI width (default is 95%)
`round.digits`	rounding digits for the output (default is 3)

Value

returns a list object containing some or all of the following:

`D`	Mahalanobis' *D*
`CI_exact_D`	exact CI for *D* (lower bound, upper bound); NA if not solvable
`OVL`	overlapping coefficient *OVL* (single distribution)
`OVL2`	Cohen's coefficient of overlap OVL_2 (joint distribution; equals 1-U_1)
`CL`	common language effect size *CL*
`PCC`	probability of correct classification (*PCC*; linear discriminant analysis)
`H2`	heterogeneity coefficient H_2
`EPV2`	EPV_2 coefficient (equivalent proportion of variables)
`CC_cor`	Tucker's Congruence Coefficient *CC* (similarity between correlation matrices)
`steiger.p`	*p* value for Steiger's test of equality of correlation matrices (null hypothesis: equality)
`d_values`	vector of Cohen's *d* values
`Du`	bias-corrected D_u
`OVLu`	overlapping coefficient (single distribution) based on D_u
`OVL2u`	Cohen's coefficient of overlap (joint distribution) based on D_u
`CLu`	common language effect size based on D_u
`PCCu`	probability of correct classification based on D_u
`du_values`	vector of bias-corrected d_u values (equivalent to Hedges' *g*)
`Dc`	disattenuated D_c
`OVLc`	disattenuated overlapping coefficient (single distribution)
`OVL2c`	disattenuated Cohen's coefficient of overlap (joint distribution)
`CLc`	disattenuated common language effect size
`PCCc`	disattenuated probability of correct classification
`H2c`	disattenuated heterogeneity coefficient H_2
`EPV2c`	disattenuated EPV_2 coefficient
`dc_values`	vector of disattenuated Cohen's d_c values
`Dcu`	disattenuated, bias-corrected D_{cu}
`OVLcu`	disattenuated overlapping coefficient (single distribution) based on D_{cu}
`OVL2cu`	disattenuated Cohen's coefficient of overlap (joint distribution) based on D_{cu}
`CLcu`	disattenuated common language effect size based on D_{cu}
`PCCcu`	disattenuated probability of correct classification based on D_{cu}
`dcu_values`	vector of disattenuated, bias-corrected d_{cu} values (disattenuated Hedges' *g*)