DIF Magnitude Functions by Rich - Demonstration and Validation

Rich Jones

2025-08-12

# Overview

This document demonstrates some DIF magnitude measures. The DIF magnitude measure functions are written to take advantage of model specification and output derived from Mplus analyses (e.g., H5 output), but are defined in such a way that they could be operationalized using any latent variable modeling software. The DIF magnitude measures are defined for:

* continuous items
* dichotomous items
* ordered categorical items
* total scale impact.

A common framework for expressing DIF is used, one that computes a difference in expected response functions (ERF) across the ability distribution between a reference and a focal group. The difference is integrated across the ability distribution and weighted according to the expected distribution of the underlying latent trait for the focal group. DIF effect size measures are generated in the natural scale of the response variable, and standardized to a common effect size measure. The standardized effect sizes are scaled to a metric consistent with Cohen’s *d* (for continuous response variables) and Cohen’s *h* for ordered categorical and dichotomous response variables (Cohen, 1988). Signed and unsigned (or compensating and non-compensating) summary measures are computed. Standard errors are defined for most magnitude measures, but some effort is required to obtain the relevant covarince parameters from Mplus output to obtain these estimates.

# Explanation of the computeAreas Function

This section explains the functionality of the computeAreas function, including its components, calculations, and outputs.

## 1. Unsigned Area

The **unsigned area** is always computed as the square root of the integral of squared differences across the full range of (latent trait).

**Formula**:

## 2. Signed Area

If the slopes differ (), the **signed area** is computed by calculating the Euclidean distances (unsigned areas) above and below the crossover point $ x\_c $, and subtracting them.

**Formula**:

Where: - : The crossover point between the two curves. - : The probability density function of the latent trait, defined as:

## 3. Standardized Areas

If the pooled standard deviation is supplied, the function computes standardized versions of both unsigned and signed areas.

**Formulas**:

## 4. Standard Errors

If a covariance matrix is supplied, the function uses the delta method to compute the standard errors of the unsigned and signed areas. The gradient for the unsigned area is computed with respect to the parameters . If the slopes differ, the gradient for the signed area is also computed. If a covariance matrix for the item parameters is supplied along with a standard deviation, the standard error for the standardized area is estimated.

where refers to the signed area or unsigned area.

**Table 1.** DIF Effect size functions

| Functions | Item type | Description |
| --- | --- | --- |
| computeAreas | Continuous | Difference in expected item score for focal group members, where the difference is defined as the Euclidean distance comparing focal versus reference group ERFs, weighted to the population distribution of ability in the focal group. Standardized versions use Cohen’s *d* metric. |
| B | Binary | Difference in expected proportion correct for focal group members, where the difference is as the Euclidean distance defined comparing focal versus reference group ERFs, weighted to the population distribution of ability in the focal group. Standardized versions use Cohen’s *h* metric (which is on the same scale as Cohen’s *d*) |
| A | Ordered categorical | Difference in expected item score (proportion of available item score) for focal group members, where the difference is defined as the Euclidean distance comparing focal versus reference group ERFs, weighted to the population distribution of ability in the focal group. Standardied versions use Cohen’s *h* metric. |

## Functions

### Supporting functions

* **function-mplus\_engine.r** defines a knitr engine that recognizes the Mplus executable. Allows for running Mplus code from a RMD or QMD document in R/Studio

### Functions for Continuous items

* **function-computeAreas.r** defines computeArea function, which computes the unsigned area (UA) and signed area (SA) between the expected response function for a continuous response variable, weighted to the expected (focal) group ability distribution, and standard error if a covariance matrix of parameters is provided. This statistic is analogous to Nye and Drasgow’s *dMACS* statistic for unsigned area, but computes the signed area differently (by computing Euclidean distances above and below a crossover point). Also computes the standardized versions of these statistics if a standard deviation is provided (pooled standard deviation of the response variable). Outputs of computeAreas function include:

1. unsigned\_area: Raw unsigned area.
2. signed\_area: Raw signed area (if slopes differ; otherwise NA).
3. std\_unsigned\_area: Standardized unsigned area (if pooled SD is supplied).
4. std\_signed\_area: Standardized signed area (if slopes differ and pooled SD is supplied).
5. unsigned\_se: Standard error of the unsigned area (if covariance matrix is supplied).
6. signed\_se: Standard error of the signed area (if slopes differ and covariance matrix is supplied).
7. std\_unsigned\_se: Standard error of the standardized unsigned area.
8. std\_signed\_se: Standard error of the standardized signed area.

* **function-plot\_lines.r** defines a function plot\_lines that uses the plot function to generate a graphic illustrating the divergence of two ERFs for continuous response variables. Highlights the area between the two lines over a interval of -4 to 4.
* **function-pooled\_sd.r** defines the pooled\_sd function, that will compute the pooled standard deviation of a dependent variable, across a grouping variable, in a data frame.
* **function-pull\_ItemParameters.r** defines pull\_ItemParameters that will pull intercepts (for continuous dependent variables) and measurement slopes (“BY”) parameters from a specified Mplus H5 object for two specified groups (i.e., is meant to be run after a multiple group CFA model). This function returns an object c(a1,b1,a2,b2) where a refers to an *intercept* and b refers to a *slope* of the regression of the continuous response variable on the continuous latent factor in the measurement model: . (Please note that a and b take on different roles in typically presented equations of the two parameter item response theory model, where and is the intercept and is the slope of the regression of a transformation of the response variable on the latent variable in the measurement model.
* **function-pull\_muSigma.r** defines pull\_muSigma that will pull the mean and (residual) variance of a latent variable in a specified group from an Mplus H5 object. The values are obtained from the parameter estimates, and therefore if the model includes covariances these are residual covariances and should not be used for standardization and effect size estimation: use Tech 4 output instead. (There is no function for that.)

## Notes on DIF measures for continuous items

The provided functions generate item-level effect size measures for measurement invariance (or differential item functioning) analyses with continuous response variables. The functions are created to be used with output generated from Mplus software, specifically the results saved with H5RESULTS (Mplus version 8.11+).

The function Area\_measures computes signed, unsigned, and standardized area-between-the-curves statistics, and plots the two expected response functions (ERFs) using base R. The area measures are integrated over the presumed population distribution of the ability distribution in the focal group.

## BACKGROUND

I was inspired to code them based on the work of Nambury Raju’s use of the same in the context of Differential Item Functioning in item response theory (Raju, 1988, Raju et al, 1995). Raju et al (1995) introduces the DFIT (differential functioning of items and tests) procedure, which builds on the signed area and unsigned area statistics introduced in Raju (1988), replacing them with CDIF and NCDIF (compensatory and non-compensatory differential item functioning) statistics, and their use in detecting items with DIF in the context of ordinal dependent variables. Nye and Drasgow (year) have described these statistics for continuous response variables, including their standardized versions.

There is an existing R package for doing the DFIT for ordinal dependent variables.

### Differences of our effect size statistics dUA, dSA relative to Nye & Drasgow (2011) and Gunn’s (2019) d\_MACS framework

The standardized, weighted, area measures coded here have been previously described as and d\_MACS\_signed (see Gunn et al 2020? 2019?).

There is an existing R package for computation including from mplus objects:

devtools::install\_github("ddueber/dmacs")

#### Nye et al on effect size interpretation

Nye et al (2019) suggest (and by extension, our dUA) measures of less than 0.20 are indicative of negligible non-equivalence and thresholds of 0.2, 0.4, and 0.7 can be used to identify items with small, medium and large DIF effects.

## Usage notes

### Using H5RESULTS

The user must have the environment set up to deal with the H5RESULTS from Mplus. Please see <https://www.statmodel.com/mplus-R/H5ResultsTutorial.pdf>. Doug Tommet has made a package for easy loading of these functions:

devtools::install\_github("dougtommet/mplush5")

### Mplus is CaSe InSensItiVE

CAUTION: Remember Mplus is case insensitive, meaning regardless of how data are named in the source data set (and perhaps in group\_df), the H5RESULTS will have variable names in ALLCAPS. I have written these functions so that the user can and should use ALLCAPS in the command but if the source data do not have variable names in ALLCAPS it will still work because the function renames toupper the variables in the source data.

### computeAreas demonstration

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.2 ✔ tibble 3.3.0  
✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
✔ purrr 1.1.0   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors  
here() starts at /Users/rnj/DWork/GitHub/DIFMagnitude  
  
  
Attaching package: 'janitor'  
  
  
The following objects are masked from 'package:stats':  
  
 chisq.test, fisher.test  
  
  
Version: 1.1.1  
We work hard to write this free software. Please help us get credit by citing:   
  
Hallquist, M. N. & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. Structural Equation Modeling, 25, 621-638. doi: 10.1080/10705511.2017.1402334.  
  
-- see citation("MplusAutomation").  
  
Bioconductor version '3.20' is out-of-date; the current release version '3.21'  
 is available with R version '4.5'; see https://bioconductor.org/install  
  
Bioconductor version 3.20 (BiocManager 1.30.26), R 4.4.3 (2025-02-28)

Warning: package(s) not installed when version(s) same as or greater than current; use  
 `force = TRUE` to re-install: 'rhdf5'

Old packages: 'broom', 'datawizard', 'DEoptimR', 'dials', 'future', 'GGally',  
 'glmnet', 'httr2', 'modeldata', 'parallelly', 'pbapply', 'pbkrtest',  
 'probably', 'pROC', 'quarto', 'reticulate', 'Rmpfr', 'rsample', 'SimDesign',  
 'sjmisc', 'skimr', 'statsExpressions', 'V8'

New names:  
Rows: 3130 Columns: 10  
── Column specification  
──────────────────────────────────────────────────────── Delimiter: "," dbl  
(10): ...1, vdmde1z, vdmde2z, vdmde3, vdmde4z, vdmde5z, vdmre1z, vdmre2z...  
ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
• `` -> `...1`

# A tibble: 3 × 4  
 group3 group3lab ethlang n  
 <dbl> <chr> <dbl> <int>  
1 1 NHW 3 2535  
2 2 HISPENG 1 359  
3 3 HISPSPAN 2 236

We’ll use the computeArea function to get the summary estimates of the differences between the two expected response functions for one of the items in that DIF analysis.

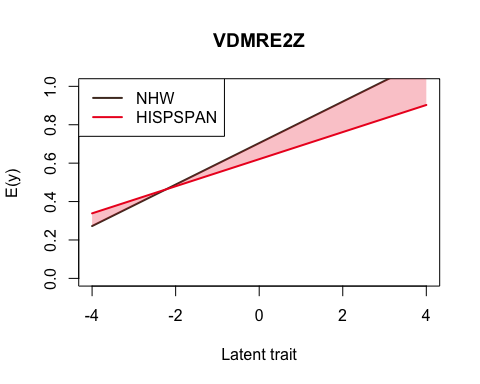
# Example usage:  
  
h5is <- here::here("DATA","mm06.h5")  
itemis <- "VDMRE2Z"  
group1is <- "NHW"  
group2is <- "HISPSPAN"  
factoris <- "GENMEM"  
groupvaris <- "group3"  
  
params <- pull\_itemParameters(h5is,group1is,group2is,itemis,factoris)  
musigma <- pull\_muSigma(h5is,group2is,factoris)  
pooled\_sd\_value <- df |> pooled\_sd(groupvaris, itemis)

[1] "Pooled sd over 3 levels of GROUP3 = 0.175015289684169"

results <- computeAreas(params, musigma, sd = pooled\_sd\_value, cov\_matrix = NULL)  
print(results)

$unsigned\_area  
[1] 0.06234433  
  
$signed\_area  
[1] 0.05945624  
  
$std\_unsigned\_area  
[1] 0.3562222  
  
$std\_signed\_area  
[1] 0.3397203  
  
$unsigned\_se  
[1] NA  
  
$signed\_se  
[1] NA  
  
$std\_unsigned\_se  
[1] NA  
  
$std\_signed\_se  
[1] NA

df |> plot\_lines(params, group1is, group2is, itemis)



## Acknowledgements

The code and descriptive text was generated with the help of ChatGPT-4o <https://chatgpt.com/c/8886f3c1-fa9f-4d53-854d-a3b5d52be582>.

## References

Cervantes, V. H. (2017). DFIT: An R package for Raju’s differential functioning of items and tests framework. Journal of statistical software, 76, 1-24.

Chalmers, R. P., Counsell, A., & Flora, D. B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. Educational and psychological measurement, 76(1), 114-140.

Gunn, H. J., Grimm, K. J., & Edwards, M. C. (2020). Evaluation of six effect size measures of measurement non-invariance for continuous outcomes. Structural Equation Modeling: A Multidisciplinary Journal, 27(4), 503-514.

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: understanding the practical importance of differences between groups. Journal of Applied Psychology, 96(5), 966.

Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2019). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. Organizational Research Methods, 22(3), 678-709.

Oshima, T., & Morris, S. (2008). Raju’s differential functioning of items and tests (DFIT). Educational Measurement: Issues and Practice, 27(3), 43-50.

Raju, N. S. (1988). The area between two item characteristic curves. Psychometrika, 53(4), 495-502.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. Applied Psychological Measurement, 19(4), 353-368.

(fin)