

## DIF — Introduction to differential item functioning

[Description](#)[Remarks and examples](#)[References](#)[Also see](#)

## Description

Differential item functioning (DIF) occurs when items that are intended to measure a latent trait are unfair, favoring one group of individuals over another. This entry provides an overview of DIF. See the following manual entries for details about the individual DIF tests, including syntax and worked examples.

`diflogistic`

Logistic regression DIF test

`difmh`

Mantel–Haenszel DIF test

## Remarks and examples

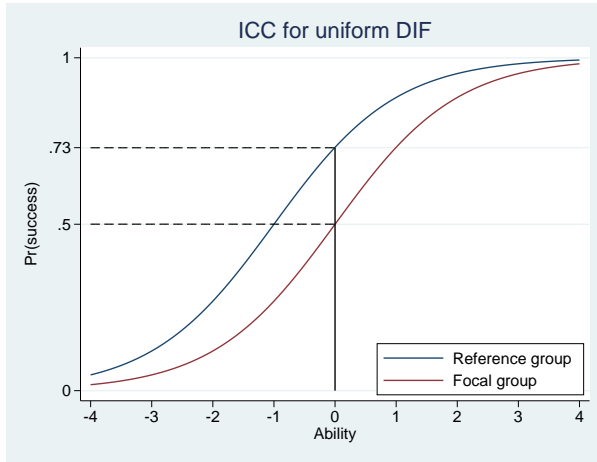
[stata.com](#)

DIF is often investigated in conjunction with fitting item response theory (IRT) models. For an introduction to the IRT features in Stata, we encourage you to read [\[IRT\] irt](#) first.

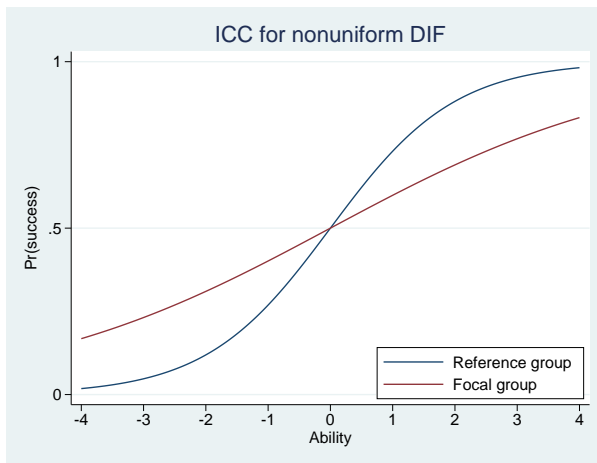
Investigating DIF involves evaluating whether a test item behaves differently across respondents with the same value of the latent trait. An item “functions differently” across individuals with the same latent trait level if these individuals have different probabilities of selecting a given response. A good overview of DIF procedures can be found in [Penfield and Camilli \(2007\)](#) and [Osterlind and Everson \(2009\)](#). [Holland and Wainer \(1993\)](#) provide a thorough treatment of statistical methodologies and practical issues surrounding DIF analysis.

It is convenient to illustrate DIF using item characteristic curves even though many DIF diagnostics are not based on fitting an IRT model. These include the Mantel–Haenszel test available in `difmh` and the logistic regression test available in `diflogistic`.

By convention, the hypothesized disadvantaged individuals are categorized as the focal group, and the advantaged ones are categorized as the reference group. The graph below shows an example of uniform DIF. The item favors the reference group over the entire range of the latent trait (ability). For example, given ability = 0, the probability of getting the item right is 0.73 for an individual from the reference group but only 0.50 for an individual from the focal group.



When the two item characteristic curves cross, we have a case of nonuniform DIF; that is to say, different groups are favored in different ranges of the latent trait. In the graph below, the item favors the focal group over the negative range of ability, and the item favors the reference group over the positive range of ability.



The MH test is formally a test of uniform DIF but can, in some cases, detect nonuniform DIF; see [IRT] [difmh](#) for details.

The logistic regression procedure can detect both uniform and nonuniform DIF; see [IRT] [diflogistic](#) for details.

## References

- Holland, P. W., and H. Wainer, ed. 1993. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.
- Osterlind, S. J., and H. T. Everson. 2009. *Differential Item Functioning*. 2nd ed. Thousand Oaks, CA: SAGE.
- Penfield, R. D., and G. Camilli. 2007. Differential item functioning and item bias. In Vol. 26 of *Handbook of Statistics: Psychometrics*, ed. C. R. Rao and S. Sinharay, 125–167. Amsterdam: Elsevier. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X).

## Also see

[IRT] [diflogistic](#) — Logistic regression DIF

[IRT] [difmh](#) — Mantel–Haenszel DIF

[IRT] [irt](#) — Introduction to IRT models