diflogistic — Logistic regression DIF

Description

diflogistic uses logistic regression to test whether an item exhibits differential item functioning (DIF) between two observed groups. Logistic regression is used to test for both uniform and nonuniform DIF, that is, whether an item favors one group over the other for all values of the latent trait or for only some values of the latent trait.

Quick start

Logistic regression test for binary items b1 to b100 using variable grpvar to identify the focal and reference groups

```
diflogistic b1-b100, group(grpvar)
```

Same as above, but request the test for items b5, b10, and b15 only

```
diflogistic b1-b100, group(grpvar) items(b5 b10 b15)
```

Replay the results, but show only items with p-values ≤ 0.05

```
diflogistic, maxp(.05)
```

Menu

Statistics > IRT (item response theory)

Syntax

diflogistic *varlist* [*if*] [*in*] [*weight*], <u>g</u>roup(*varname*) [*options*]

options	Description
Main	
* <u>g</u>roup(*varname*)	specify variable that identifies groups
total(*varname*)	specify total score variable
items(*varlist_i*)	calculate logistic regression test for items in *varlist_i* only
nolistwise	do not use listwise deletion to handle missing values
Reporting	
maxp(#)	display only items with p-value $\leq$ #
<u>sf</u>ormat(*%fmt*)	display format for χ^2 values; default is sformat(%9.2f)
<u>pf</u>ormat(*%fmt*)	display format for p-values; default is pformat(%9.4f)

* group() is required.

collect is allowed; see [U] **11.1.10 Prefix commands**.

fweights are allowed; see [U] **11.1.6 weight**.

Options

    Main

group(*varname*) specifies the numeric variable that identifies the focal group and the reference group. The groups should be coded 1 and 0, respectively. group() is required.

total(*varname*) specifies the variable to be used as a total score. By default, the total score is calculated as the row sum of the item variables.

items(*varlist_i*) requests that a logistic regression test be calculated only for the specified items. *varlist_i* must be a subset of *varlist*. By default, the statistics are calculated for all the items in *varlist*.

nolistwise specifies to omit observations where all the variables are missing. By default, observations with any missing values are omitted.

    Reporting

maxp(#) requests that only items with p-value $\leq$ # be displayed.

sformat(*%fmt*) specifies the display format used for the χ^2 values of the output table. The default is sformat(%9.2f).

pformat(*%fmt*) specifies the display format used for the p-values of the output table. The default is pformat(%9.4f).

Remarks and examples

The following discussion is about how to use `diflogistic` to test for uniform and nonuniform DIF in binary items. If you are new to DIF, we encourage you to read [IRT] **DIF** first. If you are new to item response theory or to the item response theory features in Stata, we encourage you to read [IRT] **irt** first.

▷ Example 1: Logistic regression test for uniform and nonuniform DIF

To illustrate logistic regression DIF analysis, we use an abridged version of the mathematics and science data from De Boeck and Wilson (2004). Student responses to nine test items q1–q9 are coded 1 for correct and 0 for incorrect. We want to test for DIF based on sex. Here we tabulate the variable `female`.

```
. use https://www.stata-press.com/data/r19/masc2
(Data from De Boeck & Wilson (2004))

. tabulate female
```

Female	Freq.	Percent	Cum.
Male	761	50.73	50.73
Female	739	49.27	100.00
Total	1,500	100.00	

We have 761 male and 739 female students in our dataset. The females are coded 1 and represent the focal group.

We perform the logistic regression test on items q1–q9 by typing

```
. diflogistic q1-q9, group(female)
Logistic regression DIF analysis
```

| | Nonuniform | | Uniform | |
Item	chi2	prob.	chi2	prob.
q1	1.03	0.3092	13.20	0.0003
q2	1.39	0.2388	1.80	0.1793
q3	0.39	0.5316	6.90	0.0086
q4	7.25	0.0071	4.89	0.0270
q5	2.29	0.1300	5.91	0.0150
q6	1.18	0.2780	0.43	0.5117
q7	0.04	0.8352	2.61	0.1064
q8	0.96	0.3270	2.24	0.1347
q9	0.23	0.6285	2.23	0.1352

Columns 2 and 3 report the results of logistic regression tests for the presence of nonuniform DIF. Using a 5% significance level, we conclude that only item q4 exhibits nonuniform DIF.

Columns 4 and 5 report the results of logistic regression tests for the presence of uniform DIF. These tests indicate that items q1, q3, and q5 exhibit uniform DIF, again using a 5% significance level. We use this test only for items that do not exhibit nonuniform DIF; thus, in our example, we ignore the uniform DIF test reported for item q4.

A visual examination of the output table becomes cumbersome even for a moderate number of items. We can use diflogistic to display only the items whose p-value falls below a certain significance level. Below we redisplay the results with the maxp(.05) option.

```
. diflogistic, maxp(.05)
Logistic regression DIF analysis
```

Item	Nonuniform		Uniform	
	chi2	prob.	chi2	prob.
q1	.	.	13.20	0.0003
q3	.	.	6.90	0.0086
q4	7.25	0.0071	.	.
q5	.	.	5.91	0.0150

Now it is much easier to see that item q4 exhibits a nonuniform DIF and items q1, q3, and q5 exhibit uniform DIF.

Note that neither test reports anything about the amount or direction of DIF exhibited by an item. For items that exhibit uniform DIF, you can use a common odds ratio reported by difmh to assess the amount and direction of DIF.

◁

Stored results

diflogistic stores the following in r():

Macros
    r(cmd)   diflogistic
    r(cmdline)   command as typed
    r(items)   names of items
    r(wtype)   weight type
    r(wexp)   weight expression
    r(group)   group variable
    r(total)   name of alternative total score variable, if specified

Matrices
    r(dif)   results table
    r(_N)   number of observations per item

Methods and formulas

Let L_1, L_2, and L_3 be the log-likelihood values associated with the following models, respectively,

$$\text{logit}\{\Pr(\mathbf{y})\} = \tau_0 + \tau_1 \mathbf{t} + \tau_2 \mathbf{g} + \tau_3 (\mathbf{t} \times \mathbf{g}) \tag{1}$$

$$\text{logit}\{\Pr(\mathbf{y})\} = \tau_0 + \tau_1 \mathbf{t} + \tau_2 \mathbf{g} \tag{2}$$

$$\text{logit}\{\Pr(\mathbf{y})\} = \tau_0 + \tau_1 \mathbf{t} \tag{3}$$

where $\mathbf{y}$ is a vector of responses for a given item; $\mathbf{t}$ is the latent trait, most commonly represented by the observed total score; and $\mathbf{g}$ is a dichotomous variable representing the focal group.

Likelihood-ratio tests are used to compare the nested models. The test for nonuniform DIF compares models (1) and (2) and is given by $\text{LR}_1 = -2(L_1 - L_2)$. LR_1 is distributed as χ_1^2. If the null hypothesis of no nonuniform DIF is rejected, we do not proceed to the test for uniform DIF.

The test for uniform DIF compares models (2) and (3) and is given by $\text{LR}_2 = -2(L_2 - L_3)$. LR_2 is distributed as χ_1^2.

References

De Boeck, P., and M. Wilson, eds. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer. https://doi.org/10.1007/978-1-4757-3990-9.

Swaminathan, H., and H. J. Rogers. 1990. Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement* 27: 361–370. https://doi.org/10.1111/j.1745-3984.1990.tb00754.x.

Also see

[IRT] **DIF** — Introduction to differential item functioning

[IRT] **difmh** — Mantel–Haenszel DIF

[IRT] **irt** — Introduction to IRT models