Title

difmh — Mantel–Haenszel DIF

Description   Quick start   Menu   Syntax
Options   Remarks and examples Stored results Methods and formulas
References   Also see

Description

difmh calculates the Mantel–Haenszel (MH) χ^2 and common odds ratio for dichotomously scored items. The MH statistics are used to determine whether an item exhibits uniform differential item functioning (DIF) between two observed groups, that is, whether an item favors one group relative to the other for all values of the latent trait.

Quick start

MH χ^2 and common odds ratio for binary items b1 to b100 using variable group to identify the focal and reference groups

    difmh b1-b100, group(group)

As above, but request MH statistics for items b5, b10, and b15 only

    difmh b1-b100, group(group) items(b5 b10 b15)

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

    difmh, maxp(.05)

Menu

Statistics > IRT (item response theory)

Syntax

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

options	Description
Main	
* <u>group</u>(*varname*)	specify variable that identifies groups
<u>total</u>(*varname*)	specify total score variable
<u>items</u>(*varlist$_i$*)	calculate MH statistics for items in *varlist$_i$* only
<u>noyates</u>	do not apply Yates's correction for continuity; default is to apply the continuity correction
Reporting	
<u>level</u>(*#*)	set confidence level; default is level(95)
<u>maxp</u>(*#*)	display only items with *p*-value ≤ *#*
<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)
<u>of</u>ormat(*%fmt*)	display format for odds-ratio statistics; default is oformat(%9.4f)

* group() is required.

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 MH statistics 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*.

noyates specifies that Yates's correction for continuity not be applied when calculating the MH χ^2 statistic. By default, the continuity correction is applied.

    ⌐ **Reporting** ⌐

level(*#*); see [R] estimation options.

maxp(*#*) requests that only items with *p*-value ≤ *#* 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).

oformat(*%fmt*) specifies the display format used for the odds-ratio statistics of the output table. The default is oformat(%9.4f).

Remarks and examples

The following discussion is about how to use difmh to test for uniform 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: MH test of uniform DIF

To illustrate the MH 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 http://www.stata-press.com/data/r14/masc2
(Data from De Boeck & Wilson (2004))

. tabulate female
```

1=female, 0=male	Freq.	Percent	Cum.
0	761	50.73	50.73
1	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 MH procedure on items q1–q9 by typing

```
. difmh q1-q9, group(female)
Mantel-Haenszel DIF Analysis
```

Item	Chi2	Prob.	Odds Ratio	[95% Conf. Interval]	
q1	12.47	0.0004	1.6053	1.2395	2.0790
q2	1.79	0.1813	1.1809	0.9354	1.4907
q3	6.58	0.0103	1.4543	1.0993	1.9238
q4	3.86	0.0496	0.7879	0.6241	0.9947
q5	5.00	0.0253	0.7011	0.5189	0.9472
q6	0.49	0.4835	1.1046	0.8567	1.4243
q7	1.77	0.1836	0.8359	0.6500	1.0750
q8	2.09	0.1478	0.7761	0.5615	1.0727
q9	2.03	0.1546	0.8294	0.6479	1.0618

The Chi2 and Prob. columns contain the MH χ^2 statistic with the associated significance level. Items q1, q3, q4, and q5 exhibit DIF based on a 5% significance level. However, significant statistics do not tell us anything about the amount or direction of DIF exhibited by an item.

The last three columns present the MH common odds ratio with the associated confidence interval. A common odds ratio greater than 1 indicates DIF in favor of the focal group. The results suggest that items q1 and q3 favor females, and items q4 and q5 favor males.

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

```
. difmh, maxp(.05)

Mantel-Haenszel DIF Analysis
   Item |      Chi2       Prob. | Odds Ratio      [95% Conf. Interval]
--------+----------------------+-------------------------------------
     q1 |     12.47      0.0004 |     1.6053        1.2395      2.0790
     q3 |      6.58      0.0103 |     1.4543        1.0993      1.9238
     q4 |      3.86      0.0496 |     0.7879        0.6241      0.9947
     q5 |      5.00      0.0253 |     0.7011        0.5189      0.9472
--------+----------------------+-------------------------------------
```

◁

Stored results

difmh stores the following in r():

Scalars
    r(N)   number of observations
    r(level)   significance level
    r(yates)   1 if Yates's continuity correction is used, 0 otherwise

Macros
    r(cmd)   difmh
    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(sigma2)   estimated variance of the common odds ratio
    r(_N)   number of observations per item

Methods and formulas

The MH test, also known as the Cochran–Mantel–Haenszel test, is used to determine whether two dichotomous variables are independent of one another after conditioning on a third variable; see Mantel and Haenszel (1959) and Holland and Thayer (1988) for details.

In item response theory, one dichotomous variable represents the reference and the focal group, and the other represents a response to an item scored as correct and incorrect. The conditioning variable is the latent trait, most commonly represented by the observed total score. For a dichotomously scored instrument of length K, the total score ranges from 0 to K.

The MH χ^2 statistic is based on the sum of the 2×2 contingency tables calculated for each value (stratum) of the total score. A single 2×2 table for the kth score is

	Correct	Incorrect	Total
Reference	n_{11k}	n_{12k}	$n_{1.k}$
Focal	n_{21k}	n_{22k}	$n_{2.k}$
Total	$n_{.1k}$	$n_{.2k}$	n_k

Incomplete contingency tables do not contribute to the MH statistic; this includes perfect scores and zero scores. For items with fewer than two complete contingency tables, difmh reports missing values.

The MH χ^2 statistic is given as

$$\mathrm{MH}_{\chi^2} = \frac{\left[\left|\sum_{k=1}^{K-1}\{n_{11k} - E(n_{11k})\}\right| - c\right]^2}{\sum_{k=1}^{K-1}\mathrm{var}(n_{11k})}$$

where

$$E(n_{11k}) = \frac{n_{1.k}n_{.1k}}{n_k}$$

$$\mathrm{var}(n_{11k}) = \frac{n_{1.k}n_{2.k}n_{.1k}n_{.2k}}{n_k^2(n_k - 1)}$$

and $c = 0$ when option noyates is specified; $c = 1/2$ otherwise.

The statistic is evaluated against a χ^2 distribution with one degree of freedom. A significant MH χ^2 statistic suggests the presence of DIF in an item, however, the statistic does not indicate the amount of DIF.

To assess the amount of DIF in an item, we can use the MH common odds-ratio (OR) statistic. The statistic is given as

$$\mathrm{OR}_{\mathrm{MH}} = \frac{\sum_{k=1}^{K-1} n_{11k}n_{22k}/n_k}{\sum_{k=1}^{K-1} n_{12k}n_{21k}/n_k}$$

Under the null hypothesis of no DIF, $\mathrm{OR}_{\mathrm{MH}} = 1$.

The confidence interval for $\mathrm{OR}_{\mathrm{MH}}$ is based on the variance estimate of $\log(\mathrm{OR}_{\mathrm{MH}})$ proposed by Robins, Breslow, and Greenland (1986) and Phillips and Holland (1987):

$$\widehat{\sigma}^2 = \widehat{\mathrm{var}}\{\log(\mathrm{OR}_{\mathrm{MH}})\}$$

$$= \frac{\sum_{k=1}^{K-1}(n_{11k} + n_{22k})(n_{11k}n_{22k})/n_k^2}{2(\sum_{k=1}^{K-1} n_{11k}n_{22k}/n_k)^2}$$

$$= \frac{\sum_{k=1}^{K-1}\{(n_{11k} + n_{22k})(n_{12k}n_{21k}) + (n_{12k} + n_{21k})(n_{11k}n_{22k})\}/n_k^2}{2(\sum_{k=1}^{K-1} n_{11k}n_{22k}/n_k)(\sum_{k=1}^{K-1} n_{12k}n_{21k}/n_k)}$$

$$= \frac{\sum_{k=1}^{K-1}(n_{12k} + n_{21k})(n_{12k}n_{21k})/n_k^2}{2(\sum_{k=1}^{K-1} n_{12k}n_{21k}/n_k)^2}$$

The $100(1 - \alpha/2)\%$ confidence interval for $\mathrm{OR}_{\mathrm{MH}}$ is then given by

$$[\mathrm{OR}_{\mathrm{MH}} \times \exp(-z\widehat{\sigma}), \mathrm{OR}_{\mathrm{MH}} \times \exp(z\widehat{\sigma})]$$

References

De Boeck, P., and M. Wilson, ed. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach.* New York: Springer.

Holland, P. W., and D. T. Thayer. 1988. Differential item performance and the Mantel–Haenszel procedure. In *Test Validity*, ed. H. Wainer and H. I. Braun, 129–145. Hillsdale, NJ: Lawrence Erlbaum.

Mantel, N., and W. Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 22: 719–748. Reprinted in *Evolution of Epidemiologic Ideas: Annotated Readings on Concepts and Methods,* ed. S. Greenland, pp. 112–141. Newton Lower Falls, MA: Epidemiology Resources.

Phillips, A., and P. W. Holland. 1987. Estimators of the variance of the Mantel–Haenszel log-odds-ratio estimate. *Biometrics* 43: 425–431.

Robins, J. M., N. E. Breslow, and S. Greenland. 1986. Estimators of the Mantel–Haenszel variance consistent in both sparse data and large-strata limiting models. *Biometrics* 42: 311–323.

Also see

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

[IRT] **diflogistic** — Logistic regression DIF

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