Mantel-Haenszel Differential Item Functioning (DIF) Analyses Using SPSS Syntax

In this document I explain how to use SPSS syntax to examine differential item functioning (DIF) using the Mantel-Haenszel (MH) method discussed in Chapter 16.

As explained in that chapter, DIF can be uniform or nonuniform. In uniform DIF, only the difficulty parameter (intercept or threshold) of the item differs across groups.

In non-uniform DIF, the discrimination (or slope) of the item differs across groups. Of course, items may display both uniform and nonuniform DIF (differences across groups in both difficulty and discrimination).

The MH method tests only for uniform DIF. Methods based on logistic regression test for both uniform and nonuniform DIF. LR methods are described in a separate document.

The data for this example are from a 31-item multiple choice test for which gender DIF was assessed. Data were available from 500 male and 803 female students. The data are available in the file "DIF data.sav." There are no missing data. For these data, males are coded as 0 and females are coded as 1.

Basic DIF testing requires that each item be tested separately. With a large number of items, this can be tedious. Because of this, I created a mini-program called a "macro" that uses SPSS syntax to automate the process of running the analysis sequentially for all items and collating the results of these analyses.

In the rest of this document I first discuss the MH analysis for a single item. I then present and discuss the macro for the MH procedure. Finally, I present and discuss a program to combine and format the results of the macro-based analyses. Please refer to Chapter 16 for more detail on the MH procedures for detecting DIF.

Mantel-Haenszel Procedure for a Single Item

The Mantel-Haenszel procedure can be obtained from the SPSS **CROSSTABS** procedure using the syntax below in which I use the generic variable names "item," "group," and "totalscore" for the item being tested for DIF, the grouping variable, and the total score used for matching.

```
CROSSTABS
 /TABLES=item BY group BY totalscore
 /FORMAT=NOTABLES
 /STATISTICS=CMH(1)
 /COUNT ROUND CELL.
```

The subcommand **FORMAT=NOTABLES** will produce the MH odds ratio (α_{MH}) and chi-square (χ^2_{MH}) values.

The MH Odds Ratio (α_{MH})

Here I illustrate the MH procedure for Item 1. The MH odds ratio (α_{MH}) is shown in the output table below.

On page 484 of the text I explain that the odds of a correct answer are calculated as the proportion of correct answers divided by the proportion of incorrect answers for a particular group. The odds ratio is the odds for the reference group divided by the odds for the focal group.

Confusingly, however, in SPSS the odds are always calculated as the proportion for the lower category divided by the proportion for the higher category. Because incorrect answers are typically coded as 0 and correct answers as 1, in SPSS the odds are calculated in the opposite way from that in the text: as the proportion of *incorrect* answers divided by the proportion of *correct* answers, or the odds of an *incorrect* answer.

The good news is that, because SPSS provides the inverse of the value we want, we can simply invert the value obtained from SPSS to get the usual DIF α_{MH} value of correct proportion/incorrect proportion.

A related point is that, in SPSS the odds ratio α_{MH} is calculated as the odds for the group with the lower code divided by the odds for the group with the higher code. Because by convention values of α_{MH} for DIF analyses are calculated as the odds for the reference group divided by the odds of the focal group, I have coded the reference group (males) as 0 and the focal group (females) as 1.

The SPSS table below shows the output for item 1. The MH common odds ratio, or α_{MH} value, is .743. As noted previously, this is the odds of $p(in$correct$)/p($correct$)$ for the reference group (males, coded 0) divided by the corresponding odds for the focal group (females, coded 1). So, the odds of getting the item *incorrect* in the reference group (males) was .743 times less than the odds for the focal group (females).

We can invert this value by simply taking $1/\alpha_{MH}$, or 1/.743, which yields ≈ 1.35. This inverted value indicates that the odds of getting the item *correct* for the focal group (males) is 1.35 times the odds for the focal group (females). Thus, males have greater odds, or a greater chance, of getting this item right, after controlling for overall ability.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			.743
ln(Estimate)			-.296
Standard Error of ln(Estimate)			.160
Asymptotic Significance (2-sided)			.063
Asymptotic 95% Confidence Interval	Common Odds Ratio	Lower Bound	.544
		Upper Bound	1.017
	ln(Common Odds Ratio)	Lower Bound	-.609
		Upper Bound	.017

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

As noted in the text (see page 485), the value of α_{MH} is sometimes converted to the so-called delta (Δ_{MH}) metric by taking the natural log of α_{MH} and multiplying this by -2.35. Converting the inverted α_{MH} value of 1.35 to the delta metric yields $\Delta_{MH} \approx -.70$. Although higher values of α_{MH} indicate that the item is *easier* for the reference group, Δ_{MH} has a somewhat different interpretation. The delta value is zero if there is no DIF. It is positive if the item is easier for the focal group and negative if the item is easier for the reference group. Thus, values expressed in the delta metric will always be consistent with those expressed in the alpha metric in identifying the group for which the item is easier.

The MH Chi-square Value (χ^2_{MH})

Another aspect of DIF testing using Mantel-Haenszel is the MH chi-square (χ^2_{MH}) and its associated significance test. Both α_{MH} and Δ_{MH} can be thought of effect size metrics. The MH chi-square provides a statistical test of the equality of odds ratios in the reference and focal groups. Running the syntax above also yields the MH chi-square value, which is shown below for item 1.

Tests of Conditional Independence

	Chi-Squared	df	Asymptotic Significance (2-sided)
Cochran's	3.502	1	.061
Mantel-Haenszel	3.148	1	.076

The MH chi-square value of 3.148 is not statistically significant at the .05 level, indicating that the DIF is not significant. Using the ETS classification system for DIF, item 1 would be classified into level A because the absolute value of Δ_{MH} is less than 1 and χ^2_{MH} is not statistically significant. Level A items are considered to have negligible, if any, DIF.

Automating the process

As noted previously, DIF must be tested for each item, which can become tedious when there are many items on a scale. I created the macro below to automate the process in SPSS. In the subsequent text I explain how it works. To run these commands, the datafile DIF data.sav (or, more generally, the datafile containing the item-level data to be analyzed) should be open.

```
OMS
/SELECT TABLES
/IF COMMANDS = ['Crosstabs']
    SUBTYPES = ['Mantel-Haenszel Common Odds Ratio Estimate']
/DESTINATION FORMAT = SAV
OUTFILE = 'your folder\MH_coefficients.sav'.
OMS
/SELECT TABLES
/IF COMMANDS =['Crosstabs']
    SUBTYPES = ['Tests of Conditional Independence']
/DESTINATION FORMAT = SAV
OUTFILE = 'your folder\chi_coefficients.sav'.

DEFINE MHmac (!POS !TOKENS(31))
CROSSTABS
 /TABLES = !1 BY group BY total
 /FORMAT=NOTABLES
 /STATISTICS=CMH(1)
 /COUNT ROUND CELL.
 !ENDDEFINE.
MHmac I1 to I31.
OMSEND.
```

The **OMS** (Output Management System) procedure in SPSS can be used to select output from analyses and direct this to external files. In the set of **OMS** commands above, I use two OMS commands to direct specific values from two different tables that are produced from the CROSSTABS syntax introduced previously.

The subcommand **/SELECT TABLES** tells SPSS to select values from tables in the output. The specific tables from which values should be taken are indicated by the **IF COMMANDS** and **SUBTYPES** subcommands.

For the first set of **OMS** commands, the **IF COMMANDS** subcommand specifies that output under the heading "Crosstabs" should be identified. The **SUBTYPES =** subcommand specifies that only output from tables with the heading "'Mantel-Haenszel Common Odds Ratio Estimate'" should be saved (note that this is the heading for the table containing the α_{MH} value shown previously).

For the second set of **OMS** commands, the **IF COMMANDS** subcommand again specifies that output under the heading "Crosstabs" should be identified. The **SUBTYPES =**

subcommand specifies that only output from tables with the heading "Tests of Conditional Independence" should be saved (note that this is the heading for the table containing the χ^2_{MH} value shown previously).

Each set of **OMS** commands contain the subcommand **DESTINATION FORMAT = SAV**. This specifies that the output should be saved into an SPSS system file (SAV file). Finally, the **OUTFILE =** commands specify the drive locations (if needed) and filenames to which the two sets of output should be saved.

The next set of commands, beginning with **DEFINE** MHmac **(arg1 = !TOKENS(1)** contains the macro setup.

The commands **DEFINE** and **!ENDDEFINE** indicate the beginning and end of the macro commands.

The use of the "!" symbol indicates to SPSS that the command is a macro command.

DEFINE MHmac indicates that a macro named "MHmac" is being defined (names of macros are arbitrary but must follow the naming conventions for SPSS variables).

The "!" symbol also serves as a wildcard placeholder that can take on different values specified in the body of the macro. In this macro, I use "positional tokens" by including the syntax **(!POS !TOKENS(31))** after the **DEFINE** command. This indicates that the analyses specified within the macro should be carried out for the number of times specified by the token. In this case, the number 31 in parentheses after **TOKEN** indicates that there will be 31 such analyses. The 31 variables for which the analyses are to be carried out are specified in the subcommand

MHmac I1 to I31.

Note that this is the last command of the macro and must therefore end with a period. This command specifies that the subsequent crosstabs commands should be run for each variable from I1 to I31.

The crosstabs statement is also modified by the macro. The command:

 /TABLES=!1 BY group **BY** total

specifies that the positional variable names provided in the command **MHmac I1 to I31.** should be substituted in when the positional character (!1) is encountered. This will cause the entire set of crosstabs commands to be run for I1, I2, I3, etc., up to I31.

The **!ENDDEFINE** command specifies the end of the macro MHmac.

Finally, the command **OMSEND** indicates the end of the output to be sent to an external file.

Output from the macro

Running the syntax above will produce two new SPSS data files: one containing the α_{MH} values (MH_coefficients.sav) and one containing the χ^2_{MH} values (chi_coefficients.sav).

The file MH_coefficients.sav looks like this (below I display the values for only the first item):

Command_	Subtype_	Label_	Var1	Var2	Var3	Var4
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Estimate			.743
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	ln(Estimate)			-.296
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Standard Error of ln(Estimate)			.160
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Asymptotic Significance (2-sided)			.063
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Asymptotic 95% Confidence Interval	Common Odds Ratio	Lower Bound	.544
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Asymptotic 95% Confidence Interval	Common Odds Ratio	Upper Bound	1.017
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Asymptotic 95% Confidence Interval	ln(Common Odds Ratio)	Lower Bound	-.609
Crosstabs	Mantel-Haenszel Common Odds Ratio Estimate	Mantel-Haenszel Common Odds Ratio Estimate	Asymptotic 95% Confidence Interval	ln(Common Odds Ratio)	Upper Bound	.017

This file contains all eight values from the Mantel-Haenszel Common Odds Ratio Estimate table shown previously, for each of the 31 items. I include a program at the end of this document to strip off the α_{MH} values for each item and save these into a file along with the chi-square values for each item.

The file chi-coefficients.sav looks like this (below I display the values for only the first item):

	Command_	Subtype_	Label_	Var1	ChiSquared	df	AsymptoticSignificance2sid..
1	Crosstabs	Tests of Conditional Independence	Tests of Conditional Independence	Cochran's	3.502	1	.061
2	Crosstabs	Tests of Conditional Independence	Tests of Conditional Independence	Mantel-Haenszel	3.148	1	.076

This file contains the two chi-square values, their df, and *p*-values for each of the 31 items.

Reformatting and Merging the New Datafiles

Finally, I provide a syntax file below that will strip the values of interest (α_{MH} and the χ^2_{MH} value along with its *p*-value) from the new datafiles MH_coefficients.sav and Chi_coefficients.sav and merge these values into a single datafile.

I begin with the syntax to strip the desired values from MH_coefficients.sav.

GET FILE = '*your folder*\MH_coefficients.sav'.
DATASET COPY alpha.sav.
DATASET ACTIVATE alpha.sav.
FILTER OFF.
USE ALL.
SELECT IF (Var1='Estimate').
EXECUTE.
SAVE OUTFILE = '*your folder*\alpha.sav'.

The **SELECT IF** command above is used to select only the α_{MH} values from the file MH_coefficients.sav. These values are identified by the word "estimate" under Var1 in that file.

In the commands below, I use **DELETE VARS** to delete the variables that are not needed. These include all variables except Var4, which contains the α_{MH} values.

I also use the **COMPUTE** command to create the variable "item" to identify the item number for each value of α_{MH}.

Next, I rename Var4 as alpha and use the **COMPUTE** command to take the inverse of alpha (recall that by default SPSS computes the odds an incorrect rather than a correct response. Taking the inverse of the resulting alpha value produces the odds of a correct response.)

The second **COMPUTE** statement calculates the value of delta (Δ_{MH}).

Finally, I save the new datafile as "alpha.sav."

DELETE VARS Command_ Subtype_ Label_ Var1 Var2 Var3.
COMPUTE item **= $casenum.**
RENAME VARIABLES (Var4=alpha**).**
COMPUTE alpha_inv = 1/alpha.
COMPUTE delta = ln(alpha_inv) * -2.35.
EXECUTE.
SAVE OUTFILE = '*your folder***\alpha.sav'.**

I use similar commands to strip the chi-square values and their associated *p*-values from the file chi_coefficients.sav.

In the syntax below I use the **SELECT IF** command to select only the χ^2_{MH} and associated *p*-values values from the file chi_coefficients.sav. These values are identified by the word "Mantel-Haenszel" under Var1 in that file.

GET FILE = '*your folder***\chi_coefficients.sav'.**
DATASET COPY Chi.sav**.**
DATASET ACTIVATE Chi.sav**.**
FILTER OFF.
USE ALL.
SELECT IF (Var1='Mantel-Haenszel'**).**
EXECUTE.

I then use the **DELETE VARS** command to delete the variables that are not needed. These include all variables except those labeled "Chisquared" and "AsymptoticSignificance2sided=chi_p", which contain the χ^2_{MH} and *p*-values.

I also use the **COMPUTE** command to create the variable "item" that identifies the item number for each value of χ^2_{MH}.

Finally, I rename the two variables and save the new datafile as "chi.sav.".

DELETE VARS Command_ Subtype_ Label_ Var1 df**.**
COMPUTE item **=** $casenum**.**
RENAME VARIABLES (Chisquared = chi) (AsymptoticSignificance2sided=chi_p**).**
EXECUTE.
SAVE OUTFILE = '*your folder*chi.sav'.

The final step is to merge the two new datafiles alpha.sav and chi.sav into a new file named MH.sav. The syntax below accomplishes this:

MATCH FILES
  /FILE=chi.sav
  /FILE=alpha.sav
  /BY item**.**
EXECUTE.
SAVE OUTFILE = '*your folder***MH.sav'**
  /KEEP = item chi chi_p alpha alpha_inv delta**.**

The **ADD FILES** command below is used to reorder the variables in the merged datafile so that the item number is shown in the first column.

ADD FILES file */KEEP item all.
EXECUTE.

The resulting merged datafile looks like this (here, I show only the first few rows):

	item	chi	chi_p	alpha	alpha_inv	delta
1	1.00	3.148	.076	.743	1.35	-.70
2	2.00	.536	.464	1.107	.90	.24
3	3.00	15.235	.000	1.688	.59	1.23
4	4.00	.357	.550	.917	1.09	-.20
5	5.00	6.533	.011	.689	1.45	-.87
6	6.00	.600	.439	1.112	.90	.25
7	7.00	7.033	.008	.672	1.49	-.94